Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP04/053594

International filing date: 17 December 2004 (17.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/531,866
Filing date: 22 December 2003 (22.12.2003)

Date of receipt at the International Bureau: 25 February 2005 (25.02.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

PCT/EP2004/053594

PA 1265238

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

January 04, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/531,866***
FILING DATE: ***December 22, 2003***

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. L. JACKSON
Certifying Officer

Please type a plus sign (+) inside this box → [+]

PTO/SB/16 (02-01)
Approved for use through10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL 995743924

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Valerie	Frankard	Bruxells, Begium
Vladimir	Mironov	Sint-Amandsberg, Belgium
Ana Isabel	Sanz Molinero	Gentbrugge, Belgium

☐ *Additional inventors are being named on the ___ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (280 characters max)

PLANTS HAVING MODIFIED GROWTH CHARACTERISTICS AND METHOD FOR MAKING THE SAME

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☐ Customer Number [ ] → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☑ Firm or Individual Name	Ann Pokalsky, Esq.				
Address	DILWORTH & BARRESE, LLP				
Address	333 Earle Ovington Blvd.				
City	Uniondale	State	New York	ZIP	11553
Country	US	Telephone	(516) 228-8484	Fax	(516) 228-8516

ENCLOSED APPLICATION PARTS *(check all that apply)*

☑ Specification *Number of Pages* 48

☑ Drawing(s) *Number of Sheets* 9

☐ Application Data Sheet. See 37 CFR 1.76

☐ CD(s), Number

☑ Other (specify) Sequence Listing- (8 sheets)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☑ Applicant claims small entity status. See 37 CFR 1.27.

☑ A check or money order is enclosed to cover the filing fees

☑ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 04-1121

☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): 80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☐ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE Ann R. Pokalsky

TYPED or PRINTED NAME Ann Pokalsky

TELEPHONE 516-228-8484

Date 12/22/03

REGISTRATION NO. *(if appropriate)* 34,697

Docket Number: 1187-30

CERTIFICATION UNDER 37 C.F.R. § 1.10

I hereby certify that this correspondence and the documents referred to as enclosed are being deposited with the United States Postal Service on date below in an envelope as "Express Mail Post Office to Addressee" Mail Label Number EL 995743924 US addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Dated: December 22, 2003

Ann R. Pokalsky

Ann Pokalsky

Plants having modified growth characteristics and method for making the same

The present invention concerns a method for modifying plant growth characteristics. More specifically, the present invention concerns a method for modifying plant growth characteristics by modulating expression of a nucleic acid encoding a GRUBX protein and/or by modulating activity and/or levels of a GRUBX protein in a plant. The present invention also concerns plants having modulated expression of a nucleic acid encoding a GRUBX protein and/or modulated activity and/or levels of a GRUBX protein, which plants have modified growth characteristics relative to corresponding wild type plants.

Given the ever-increasing world population, and the dwindling area of land available for agriculture, it remains a major goal of agricultural research to improve the efficiency of agriculture and to increase the diversity of plants in horticulture. Conventional means for crop and horticultural improvements utilise selective breeding techniques to identify plants having desirable characteristics. However, such selective breeding techniques have several drawbacks, namely that these techniques are typically labour intensive and result in plants that often contain heterogeneous genetic components that may not always result in the desirable trait being passed on from parent plants. Furthermore, suitable donor species for providing a desired trait may be scarce. Advances in molecular biology have allowed mankind to manipulate the germplasm of animals and plants. Genetic engineering of plants entails the isolation and manipulation of genetic material (typically in the form of DNA or RNA) and the subsequent introduction of that genetic material into a plant. Such technology has led to the development of plants having various improved economic, agronomic or horticultural traits. Traits of particular economic interest are growth characteristics such as high yield. Yield is normally defined as the measurable produce of economic value from a crop. This may be defined in terms of quantity and/or quality. Crop yield is influenced by the typical stresses to which plants or crops are subjected. Such stresses include environmental (abiotic) stresses (such as temperature stresses caused by atypical high or low temperatures; stresses caused by nutrient deficiency; stresses caused by lack or excess of water (drought, flooding), stress caused by chemicals such as fertilisers or insecticides) and biotic stresses (which can be imposed on plants by other plants (weeds, or effects of high density planting), animal pests (including stress of grazing) and pathogens). Crop yield may not only be increased by combating one or more of the stresses to which the crop or plant is subjected, but may also be

increased by modifying the inherent growth mechanisms of a plant. The inherent growth mechanisms of a plant are controlled at several levels and by various metabolic processes.

Ubiquitination refers to a modification of proteins by conjugation to ubiquitin molecules. The term ubiquitination is often extended to processes that mediate the binding of the ubiquitin protein or of proteins that mimic the ubiquitin function. It is a versatile tool for eukaryotic cells to control stability, function and subcelullar localisation of proteins. This mechanism plays a central role in protein degradation, cell cycle control, stress response, DNA repair, signal transduction, transcriptional regulation and vesicular trafficking.

Since this mechanism is at the basis of many cellular processes, ubiquitination is highly regulated and requires high substrate specificity and the existence of ample diversity in downstream effectors. A large number of ubiquitin binding proteins has been found so far. Often these proteins have a modular domain architecture. For example ubiquitin-binding proteins typically combine a ubiquitin binding domain with a variable effector domain. Others do not contain a ubiquitin binding domain but rather have tertiary structures similar to ubiquitin that can mimic certain aspects of ubiquitination (ubiquitin-like domains).

The number of ubiquitin related motifs and domains present in ubiquitin and ubiquitin-like proteins is growing as more information on genome sequences becomes available. Some prototypes of those domains are for example UBA, UBD, UIM and UBX (see for example the Pfam database (http://www.sanger.ac.uk/Software/Pfam/); Bateman et al., Nucleic Acids Research 30(1):276-280 (2002)).

The UBX domain is a module of approximately 80 amino acid residues long and of unknown function, and is present in proteins of many organisms. Most of these proteins belong to one of five evolutionarily conserved families exemplified by the human FAF1, p47, Y33K, REP8, and UBXD1 proteins (Buchberger et al. (2001) J. Mol. Biol. 307, 17-24, Carim-Todd et al. (2001) Biochim. Biophys. Acta 1517, 298-301). Typically the UBX domain is situated at the C-terminus of a protein.

FAF1, which combines a UBA domain, two putative ubiquitin-like repeats and a UBX domain, interacts with the apoptotic signalling receptor Fas (Apo-1/CD95) and, when expressed in mammalians cells, it potentiates Fas mediated apoptosis. The UBX domain of FAF1 is located in the carboxy-terminal part, and though its function is not fully characterized, it has been shown that, the fas mediated induced apoptosis requires the UBX domain, but not the Fas-binding domain in FAF1.

Rep 8, in which no other domains are recognizable besides the UBX, is present only in mammals and is proposed to have a role in reproduction due to its abundant expression in reproductive organs (Yamabe et al. (1997) Genomics 39, 198-204; Fröhlich et al. (1998) J. Cell Sci. 111, 2353-2363).

Y33K has also a modular structure with a UBA domain at the N-terminus, a central conserved region, and a UBX domain at the C-terminus. Here too, its function is unknown.

p47 is functionally best characterized. In rats it has been shown to act as a cofactor of the AAA ATPase p97 that is involved in post-mitotic reassembly of the Golgi cisternae (Kondo et al. (1997) Nature 388, 75-78). p47 interacts with the AAA ATPase through its UBX domain, which complex is postulated to act in dissociation of other protein complexes (Yuan et al. (2001) J. Mol. Biol. 311, 255-263). Furthermore, p47 homologues are found fused to the N-termini of apparently unrelated proteins in *Arabidopsis* and *Drosophila*. These fusion proteins could represent the functional linking of two proteins that otherwise perform unrelated functions (Buchberger et al., 2001). Examples include p47 linked to PP2A or to members of the prominin family.

UBXD1 is another UBX domain and has been recently described in humans and mouse (Carim-Todd et al 2001). It was found mainly expressed in reproductive tissue, but a specific function could not be assigned. Counterparts of these proteins are also found in *Arabidopsis* (41% similarity to the human protein), *Caenorhabditis elegans* and in *Drosophila melanogaster*.

It should be noted that to date no clear role, apart from that of p47, has been assigned to UBX containing proteins. Although structural evidence suggests a function of the UBX domain in ubiquitin-related processes, experimental evidence for this role is still lacking and at present it cannot be excluded that the UBX domain has a function unrelated to that of ubiquitin.

The regulation of UBX containing proteins is yet not well understood. Proteins comprising UBX domains are usually predicted to be present mainly in the cytoplasm. However, other subcellular localizations have also been reported. Thus, phosphorylation, which is a specific protein modification used to regulate activity of many proteins, has been shown to also influence transport into the nucleus of FAF-1 (Olsen et al. (2003) FEBS Lett. 546, 218-222.).

In summary, it has been proposed that animal UBX containing proteins might be involved in enhanced expression of genes related to apoptosis, cell cycling or targeting of proteins for degradation.

In *Arabidopsis*, a plant for which the genome has been fully sequenced, there are at least 20 UBX containing proteins. They can be classified according to sequence similarity in four groups, only the group corresponding to REP8 seems not to be present in plants (see Figure 1). Similar to the animal kingdom, the Ubx domains in plant proteins are present in combination with other domains like for example SEP, G6PD, PUG, or zinc fingers. UBX containing proteins and the domain structure of these proteins have been described (Buchberger (2002) Trends Cell Biol. 12, 216-221) and can be searched for in specialised

databases e.g. SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2002) Nucleic Acids Res 30, 242-244; http://smart.embl-heidelberg.de/).

PUG domains (in Peptide:N-Glycanases and other putative nuclear UBX-domain-containing proteins; Doerks et al. (2002) Genome Research 12, 47-56) co-occur in proteins with domains that are central to ubiquitin-mediated proteolysis, including UBX (in mammals and plants), UBA (in plants) and UBC domains (in *Plasmodium*). PUG containing proteins such as PNGases are believed to play a role in the unfolded protein response, an endoplasmatic reticulum (ER) quality control surveillance system that distinguishes aberrant proteins from correctly folded proteins. In some cases, it has been shown that these misfolded and/or unfolded proteins are degraded by a so-called ER-associated degradation mechanism, which involves the ubiquitin-proteasome system (Suzuki et al. (2000) J. Cell Biol. 149, 1039-1052).

Divergent forms of PUG domains are also present on kinases of the IRE1p type which are known to function in the initial stages of the unfolded protein response (Shamu and Walter (1996) EMBO J. 15, 3028-3039).

Functions of plant UBX proteins have not yet been experimentally described. Based on sequence and protein structure conservation, it is reasonable to assume that the plant proteins may perform functions similar to the homologues in other organisms. However the functions of the family of UBX proteins are still largely unknown. The contents of the references mentioned above are incorporated by reference herein as if fully set forth.

It has now been found that modulating expression of a nucleic acid encoding a Growth Related UBX domain comprising protein (hereafter named "GRUBX") in a plant gives rise to plants having modified growth characteristics. Therefore according to a first embodiment of the present invention there is provided a method for modifying the growth characteristics of a plant, comprising modulating expression in a plant of a nucleic acid encoding a GRUBX protein and/or modulating activity and/or levels in a plant of a GRUBX protein. The GRUBX gene and protein sequence (SEQ ID NO 1 & 2) were previously disclosed in WO 03/085115.

Modulating (enhancing or decreasing) expression of a nucleic acid encoding a GRUBX protein or modulation of the activity and/or levels of the GRUBX protein itself encompasses altered expression of a gene and/or altered activity and/or levels of a gene product, namely a polypeptide, in specific cells or tissues. Altered expression of a gene and/or altered activity and/or levels of a gene product may be effected, for example by chemical means and/or recombinant means. Modulating expression of a gene and/or levels of a gene product and/or modulating activity of a gene product may be effected directly through the modulation of expression of a GRUBX encoding gene and/or directly through the modulation of the activity and/or levels of a GRUBX protein. The modulated expression may result from altered

expression levels of an endogenous *GRUBX* gene and/or may result from altered expression of a GRUBX encoding nucleic acid that was previously introduced into a plant. Similarly, modulated levels and/or activity of a GRUBX protein may be the result of altered expression levels of an endogenous *GRUBX* gene and/or may result from altered expression of a GRUBX encoding nucleic acid that was previously introduced into a plant. Additionally or alternatively, the modulation of expression as mentioned above is effected in an indirect way, for example may be effected as a result of decreased or increased levels and or activity of factors that control the expression of a *GRUBX* gene or that influence the activity and/or levels and/or subcellular or tissue localisation of the GRUBX protein. In the methods of the present invention, increased expression of a nucleic acid encoding a GRUBX protein and/or increased levels of a GRUBX protein and/or increased activity of a GRUBX protein is envisaged.

Advantageously, modulation of expression of a nucleic acid encoding a GRUBX protein and/or modulation of activity and/or levels of the GRUBX protein itself may be effected by chemical means, i.e. by exogenous application of one or more compounds or elements capable of modulating activity and/or levels of the GRUBX protein and/or capable of modulating expression of a *GRUBX* gene (which may be either an endogenous gene or a transgene introduced into a plant). The term "exogenous application" taken in its broadest context includes contacting or administering cells, tissues, organs or organisms with a suitable compound or element. The compound may be applied to a plant in a suitable form for uptake (such as through application to the soil for uptake via the roots, or by applying directly to the leaves, for example by spraying).

Suitable compounds or elements for exogenous application include GRUBX encoding nucleic acids and nucleic acids that hybridise therewith; the GRUBX gene product or a homologue, derivative or active fragment thereof and/or to antibodies recognizing or mimicking the gene product. Such antibodies may comprise "plantibodies", single chain antibodies, IgG antibodies and heavy chain antibodies from camels or other members of the Camelidae, as well as fragments thereof.

Additionally or alternatively, contacting or administering cells, tissues, organs or organisms with an interacting protein or with an inhibitor or activator of the gene/gene product provides another exogenous means for modulation of expression of a nucleic acid encoding a GRUBX protein and/or for modulation of activity and/or level of the GRUBX protein itself. Modulation of expression of a nucleic acid encoding a GRUBX protein and/or modulation of activity and/or levels of the GRUBX protein itself may also be effected as a result of altered levels of factors that directly or indirectly activate or inactivate a GRUBX protein.

Plants, seeds or other plant material can also be subjected to treatment with mutagenic substances. Chemical substances effecting mutagenesis comprise N-nitroso-N-ethylurea, ethylene imine, ethyl methanesulphonate or diethyl sulphate. As an alternative, ionising radiation such as γ-rays or X-rays can equally well be used. Methods for introducing mutations and testing the effect of mutations (such as modified protein expression and/or modified protein activity) are known in the art. Therefore there is provided a method for modifying plant growth characteristics, comprising modulating expression in a plant of a nucleic acid sequence encoding an GRUBX protein and/or modulating activity in a plant of a GRUBX protein, wherein said modulation is effected by mutagenesis. Encompassed by mutagenesis are methods employing chemical mutagens, as well as physical mutagens, such as radiation.
Any characteristic of the GRUBX protein can be altered by mutagenesis. For example, the expression level can be increased or decreased, the activity of the protein can be modified, or the affinity for its interacting proteins can be adapted. According to a preferred aspect of the invention said mutagenesis results in increased expression and/or activity of a GRUBX protein.

Therefore, according to one aspect of the present invention, there is provided a method for modifying the growth characteristics of a plant and/or a method for the production of plants or plant parts having modified growth characteristics, comprising exogenous application of one or more compounds or elements capable of modulating expression of a *GRUBX* gene and/or capable of modulating activity of a GRUBX protein.

Additionally or alternatively, and according to a preferred embodiment of the present invention, modulation of expression of a nucleic acid encoding a GRUBX protein and/or modulation of activity and/or levels of the GRUBX protein itself may be effected by recombinant means. Such recombinant means may comprise a direct and/or indirect approach for modulation of expression of a nucleic acid and/or for modulation of the activity and/or levels of a protein.

For example, an indirect recombinant approach may comprise introducing into a plant, a nucleic acid capable of modulating activity and/or levels of the protein in question (a GRUBX protein) and/or capable of modulating expression of the gene in question (a gene encoding a GRUBX protein). Examples of such nucleic acids to be introduced into a plant are nucleic acids encoding transcription factors, activators or inhibitors that bind to the promoter of the *GRUBX* gene or that interact with the GRUBX protein. Methods to test these kinds of interaction and to isolate the nucleic acids encoding these interactors are for example yeast one-hybrid or yeast two-hybrid screening.

Also encompassed by an indirect approach for modulating activity and/or levels of a GRUBX protein and/or expression of a *GRUBX* gene is the inhibition or stimulation of regulatory sequences, or the provision of new regulatory sequences, that drive expression of the native gene encoding a GRUBX or the transgene encoding a GRUBX. Such regulatory sequences may be introduced into a plant. For example, the regulatory sequence introduced into the plant is a promoter, capable of driving the expression of an endogenous *GRUBX* gene.

Furthermore, modulation of the activity of a GRUBX may be effected by altering levels in a plant of a factor able to interact with GRUBX. Such factors may include ligands (regulators, subunits, substrates or targets) of the GRUBX protein.

The *GRUBX* gene or the GRUBX protein may be wild type, i.e. the native or endogenous nucleic acid or polypeptide. The nucleic acid may be derived (either directly or indirectly (if subsequently modified)) from any source provided that the nucleic acid, when expressed in a plant, leads to modulated expression of a *GRUBX* nucleic acid/gene or modulated activity and/or levels of a GRUBX protein. The nucleic acid may be isolated from a microbial source, such as bacteria, yeast or fungi, or from a plant, algal or animal (including human) source. This nucleic acid may be substantially modified from its native form in composition and/or genomic environment through deliberate human manipulation. The nucleic acid sequence is preferably a homologous nucleic acid sequence, i.e. a structurally and/or functionally related nucleic acid sequence, preferably obtained from a plant, whether from the same plant species or different.

Therefore, according to another feature of the present invention, there is provided a method for modifying the growth characteristics of a plant comprising introduction into a plant of a nucleic acid capable of modulating activity and/or levels of a GRUBX protein and/or capable of modulating expression of a *GRUBX* gene, which nucleic acid is selected from:

(i) Nucleic acids encoding transcription factors, activators of a GRUBX protein;

(ii) *GRUBX* regulatory sequences; and

(iii) Nucleic acids encoding ligands of a GRUBX protein.

A direct and preferred approach for modulating expression of a *GRUBX* gene, or modulating the activity and/or levels of a GRUBX protein on the other hand, comprises introducing into a plant a nucleic acid sequence encoding a GRUBX protein or a homologue, derivative or active fragment thereof. The nucleic acid may be introduced into a plant by, for example, transformation. Therefore, according to a preferred aspect of the present invention, there is provided a method for modifying the growth characteristics of a plant comprising introducing a

GRUBX encoding nucleic acid into a plant, wherein the GRUBX encoding nucleic acid is preferably operably linked to a seed specific promoter.

The nucleic acid may be derived (either directly or indirectly (if subsequently modified)) from any source provided that the nucleic acid, when expressed in a plant, leads to modulated expression of a *GRUBX* nucleic acid/gene or modulated activity and/or levels of a GRUBX protein. The nucleic acid may be isolated from a microbial source, such as bacteria, yeast or fungi, or from a plant, algal or animal (including human) source. Preferably the nucleic acid is derived from a eukaryotic organism. This nucleic acid may be substantially modified from its native form in composition and/or genomic environment through deliberate human manipulation. The nucleic acid sequence is preferably a homologous nucleic acid sequence, i.e. a structurally and/or functionally related nucleic acid sequence, preferably obtained from a plant, whether from the same plant species or different. The nucleic acid sequence may be isolated from a dicotyledonous species, preferably from the family Solanaceae, further preferably from *Nicotiana tabacum*. More preferably, the nucleic acid is as represented by SEQ ID NO: 1 or a functional portion thereof or a nucleic acid sequence capable of hybridising therewith or is a nucleic acid encoding an amino acid represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof.

The term *GRUBX* nucleic acid/gene, as defined herein, refers to any nucleic acid encoding a GRUBX protein, or the complement thereof. Preferably the *GRUBX* nucleic acid is from plant origin, further preferably the *GRUBX* nucleic acid encodes a GRUBX protein from the family of Solanaceae, more preferably the *GRUBX* nucleic acid is a nucleic acid sequence from *Nicotiana tabacum*, most preferably the *GRUBX* nucleic acid is a nucleic acid sequence as represented by SEQ ID NO: 1 or a portion thereof, or refers to nucleic acid sequences capable of hybridising therewith, which hybridising sequences encode proteins having GRUBX protein activity, i.e. similar biological activity to that of SEQ ID NO: 1, and also refers to nucleic acids encoding an amino acid sequence represented by SEQ ID NO: 2 or homologues, derivatives or active fragments thereof.

Also this term encompasses a variant of the nucleic acid encoding a GRUBX protein due to the degeneracy of the genetic code, an allelic variant of the nucleic acid encoding a GRUBX, different splice variant of the nucleic acid encoding a GRUBX and variants that are interrupted by one or more intervening sequences.

The term GRUBX protein, as defined herein, refers to a protein related to the human UBXD1 protein (SPTrEMBL AAH07414) and/or to a protein comprising at least a UBX and a PUG

domain, and optionally also a Zinc finger domain. Preferably, the GRUBX protein is a plant protein. Further preferably, the GRUBX protein is from the family of Solanaceae, more preferably the GRUBX is a protein from *Nicotiana tabacum*, most preferably the GRUBX is a protein as represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof, which homologues, derivatives or active fragments have similar biological activity to that of SEQ ID NO: 2. However, it should be understood that GRUBX proteins from monocotyledonous plants could equally well be used in the methods of the present invention, including GRUBX proteins from *Zea mays*, *Saccharum officinarum* (SEQ ID NO 4), *Oryza sativa* (SEQ ID NO 7), *Triticum* sp., *Hordeum* sp., and *Sorghum* sp, since these sequences are related to SEQ ID NO 2 (see Figure 1b).

The GRUBX gene of *Nicotiana tabacum* was analysed with the SMART tool and was used to screen the Pfam (Version 11.0, November 2003; Bateman et al. (2002) Nucl. Acids Res. 30, 276-280) and InterPro database (Release 7.0, 22 July 2003; Mulder et al. (2003) Nucl. Acids. Res. 31, 315-318). GRUBX comprises a UBX domain (PF00789, SM00166, IPR001012) and a PUG domain (SM00580, IPR006567). The UBX domain, as defined in InterPro, is found in ubiquitin-regulatory proteins, which are members of the ubiquitination pathway, as well as a number of other proteins including FAF-1 (FAS-associated factor 1), the human Rep-8 reproduction protein and several hypothetical proteins from yeast. The function of the UBX domain is not known although the fragment of avian FAF-1 containing the UBX domain causes apoptosis of transfected cells. In *Arabidopsis* there are approximately 20 proteins with this domain. The PUG domain is found in protein kinases, N-glycanases and other nuclear proteins in eukaryotes; and is postulated to be involved in protein-protein interactions (for a review see Suzuki & Lennarz (2003) Biochem Biophys Res Commun. 302,1-5 and Biochem Biophys Res Commun. 303, 732) and in RNA binding (Doerks et al., 2002). PUG domains are often found together with UBA or UBX domains in *Arabidopsis* proteins (Doerks et al, 2002). A consensus sequence for the UBX and PUG domains, as defined in the SMART database (Software Version 4.0, sequence database update of 15 September 2003) is given in Figure 2a, Figure 2b shows the UBX and PUG domains of respectively SEQ ID NO 2 and SPTrEMBL Q9ZU93, Figure 2c shows a BLAST alignment of these 2 proteins whereas Figures 2d and 2e display an alignment between SEQ ID NO 2 and SEQ ID NO 4, and SEQ ID NO 4 and SEQ ID NO 7, respectively. Each time, the PUG and UBX domains are indicated.

Optionally, a zinc finger domain may be present in the GRUBX protein. Zinc finger domains, as defined in InterPro, are nucleic acid-binding protein structures that were first identified in the *Xenopus laevis* transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid

residues including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The 12 residues separating the second Cys and the first His are mainly polar and basic, indicating that this region is involved in nucleic acid binding. The zinc finger motif is an unusually small, self-folding domain in which Zn is a crucial component of its tertiary structure. All Zinc finger domains bind an atom of Zn in a tetrahedral array resulting in the formation of a finger-like projection, which may interact with nucleotides in the major groove of the nucleic acid. The Zn binds to the conserved Cys and His residues. Fingers have been found to bind to about 5 base pairs of nucleic acid containing short runs of guanine residues. They have the ability to bind to both RNA and DNA, a property that has not been demonstrated yet for the helix-turn-helix motif. The zinc finger may thus represent the original nucleic acid binding protein. It has also been suggested that a Zn-centred domain could be used in a protein interaction, e.g. in protein kinase C. Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the spatial positioning of the zinc atom. In the first class to be characterized, called C2H2 (IPR007087), the first pair of zinc coordinating residues consists of cysteines, while the second pair are histidines.

Another Zinc finger domain (IPR006642) may be of the type found in the Saccharomyces cerevisiae protein Rad18. Here too, the zinc finger domain is a potential nucleic acid binding and a putative nucleotide binding sequence. The optional Zinc finger domain in the GRUBX protein is however not restricted to the C2H2 or Rad18 type, but can be any type of Zinc finger domain.

Advantageously, the method according to the present invention may also be practised using portions of a sequence represented by SEQ ID NO: 1 or by using sequences that hybridise preferably under stringent conditions to SEQ ID NO: 1 (which hybridising sequences encode proteins having GRUBX activity), or by using homologues, derivatives or active fragments of a sequence according to SEQ ID NO: 2 or by using the nucleic acids encoding these homologues, derivatives or active fragments.

Homologues of SEQ ID NO 1 may be found in various eukaryotic organisms. The closest homologues are generally found in the plant kingdom. The *Arabidopsis thaliana* genome seems to have only one GRUBX homologue (SPTrEMBL Q9ZU93 & Q8LGE5, MIPS No. At2G01650, GenBank AY084317 & AAM60904), other suitable homologues of SEQ ID NO: 2 include SEQ ID NO 4 from *Saccharum officinarum*, encoded by a nucleic acid represented in SEQ ID NO3, SEQ ID NO 7 (encoded by the nucleic acid sequence presented in SEQ ID NO 6) from *Oryza sativa*, and GenBank Accession Nos. BQ198347 and BF778922 from *Pinus taeda*.

Methods for the search and identification of GRUBX homologues would be well within the realm of persons skilled in the art. Such methods comprise comparison of the sequences represented by SEQ ID NO 1 or 2, in a computer readable format, with sequences that are available in public databases such as MIPS (http://mips.gsf.de/), GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) or EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/index.html), using algorithms well known in the art for the alignment or comparison of sequences, such as GAP (Needleman and Wunsch, J. Mol. Biol. 48, 443-453 (1970)), BESTFIT (using the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics 2, 482-489 (1981))), BLAST (Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., J. Mol. Biol. 215, 403-410 (1990)), FASTA and TFASTA (W. R. Pearson and D. J. Lipman Proc.Natl.Acad.Sci. USA 85, 2444-2448 (1988)). The software for performing BLAST analysis is publicly available through the National Centre for Biotechnology Information. The abovementioned homologues were identified using blast default parameters (for example BLASTN Program Advanced Options: G-Cost (to open a gap)=5; E-Cost (to extend a gap)=2; q-Penalty (for a mismatch)=-3; r-Reward (for a match)=1; e-Expectation value (E)=10.0; W-Word size=11; TBLASTN Program Advanced Options: G-Cost (to open a gap)=11; E-Cost (to extend a gap)=1; e-Expectation value (E)=10.0; W-Word size=3). As more genomes are being sequenced, it is expected that many more GRUBX homologues shall be identifiable.

The sequence represented by SEQ ID NO: 6 was hitherto unknown as a GRUBX encoding gene. There is therefore provided an isolated nucleic acid sequence comprising:

(i) a nucleic acid sequence represented by SEQ ID NO: 6, or the complement strand thereof;

(ii) a nucleic acid sequence encoding an amino acid sequence represented by SEQ ID NO: 7, or homologues, derivatives or active fragments thereof;

(iii) a nucleic acid sequence capable of hybridising (preferably under stringent conditions) with a nucleic acid sequence of (i) or (ii) above, which hybridising sequence preferably encodes a protein having [name protein] activity;

(iv) a nucleic acid sequence according to (i) to (iii) above which is degenerate as a result of the genetic code;

(v) a nucleic acid which is an allelic variant to the nucleic acid sequences according to (i) to (iv);

(vi) a nucleic acid which is an alternative splice variant to the nucleic acid sequences according to (i) to (v);

(vii) a nucleic acid sequence which has 75,00%, 80,00%, 85,00%, 90,00%, 95,00%, 96,00%, 97,00%, 98,00% or 99,00% sequence identity to any one or more of the sequence defined in (i) to (vi);

(viii) a portion of a nucleic acid sequence according to any of (i) to (vii) above, which portion preferably encodes a protein having GRUBX activity,

The sequence represented by SEQ ID NO: 4 was assembled from 4 EST sequences (CA154270, CA144028, BQ535511 & CA184742) and was hitherto unknown as a GRUBX protein. There is therefore provided an isolated GRUBX protein comprising at least part of one of the polypeptides selected from the group consisting of:

(i) a polypeptide as given in SEQ ID NO 4,

(ii) a polypeptide as given in SEQ ID NO 7

(iii) a polypeptide with an amino acid sequence which has at least 40.00% sequence identity, preferably 50.00%, 60.00%, 70.00% sequence identity, more preferable 80% or 90% sequence identity, most preferable 95.00%, 96.00%, 97.00%, 98.00% or 99.00% sequence identity to the amino acid sequence as given in SEQ ID NO 4 or 7,

(iv) a polypeptide comprising at least UBX domain and a PUG domain, and optionally a Zinc finger domain;

(v) a homologue, a derivative, an immunologically active and/or functional fragment of a protein as defined in any of (i) to (iv);

with the proviso that the protein sequence is not a sequence as represented by SEQ ID NO 2, or database entries Q9ZU93, AAR01744, Q9D7L9, Q9BZV1, Q99PL6, ENSANGP00000020442, Q7SXA8, Q9V8K8, Q96IK9, ENSRNOP00000037228, or AAH07414.

The term GRUBX includes proteins homologous to the GRUBX as presented in SEQ ID NO 2. Accordingly, preferred homologues to be used in the methods of the present invention comprise a UBX and a PUG domain. "Homologues" of a GRUBX protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived. To produce such homologues, amino acids of the protein may be replaced by other amino acids having similar properties (such as similar hydrophobicity, hydrophilicity, antigenicity, propensity to form or break α-helical structures or β-sheet structures). Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company).

The homologues useful in the method according to the invention have at least 40.00% sequence identity or similarity (functional identity) to the unmodified protein, alternatively at least 50.00% sequence identity or similarity to an unmodified protein, alternatively at least 60.00% sequence identity or similarity to an unmodified protein, alternatively at least 70% sequence identity or similarity to an unmodified protein. Typically, the homologues have at least 80% sequence identity or similarity to an unmodified protein, preferably at least 85% sequence identity or similarity, further preferably at least 90% sequence identity or similarity to an unmodified protein, most preferably at least 95%, 96%, 97%, 98% or 99% sequence identity or similarity to an unmodified protein. The percentage of identity can be calculated using alignment programs such as GAP.

Homologous proteins can be grouped in "protein families". A protein family can be defined by functional and sequence similarity analysis, such as, for example, Clustal W. A neighbour-joining tree of the proteins homologous to GRUBX can be generated by the Clustal W program and gives a good overview of its structural and ancestral relationship (see for example Figure 1a and b, constructed with Vector NTI Suite 5.5, Informax). In a particular embodiment of the present invention, the GRUBX homologue(s) belong(s) to the same protein family as the protein corresponding to SEQ ID NO 2.

In the *Arabidopsis* genome a family member of the GRUBX protein was identified (Q9ZU93, GenBank Refseq NM_126226). Also in other plants such as rice, sugarcane or other monocotyledonous plants, family members of the GRUBX protein were identified as shown above. Advantageously also these family members are useful in the methods of the present invention.

Two special forms of homology, orthologous and paralogous, are evolutionary concepts used to describe ancestral relationships of genes. The term "paralogous" relates to homologous genes that result from one or more gene duplications within the genome of a species. The term "orthologous" relates to homologous genes in different organisms due to ancestral relationship of these genes. The term "homologues" as used herein also encompasses paralogues and orthologues of the proteins useful in the methods according to the invention. Orthologous genes can be identified by querying one or more gene databases with a query gene of interest, using for example the BLAST program. The highest-ranking subject genes that result from the search are then again subjected to a BLAST analysis, and only those subject genes that match again with the query gene are retained as true orthologous genes. For example, to find a rice orthologue of an *Arabidopsis thaliana* gene, one may perform a BLASTN or TBLASTX analysis on a rice database (such as (but not limited to) the *Oryza sativa* Nipponbare database available at the NCBI (http://www.ncbi.nlm.nih.gov) or the

genomic sequences of rice (cultivars indica or japonica)). In a next step, the obtained rice sequences are used in a reverse BLAST analysis using an *Arabidopsis* database. The results may be further refined when the resulting sequences are analysed with ClustalW and visualised in a neighbour joining tree. The method can be used to identify orthologues in many different species.

A further way to identify a functional orthologue within a group of related proteins is to determine the expression pattern and tissue distribution of the members of this protein family, whereby sequences present in the same tissues and with a similar expression pattern are expected to perform related functions. A yet further way to identify functional homologues of a protein could be by identifying sequences with a similar conserved domain structure. Proteins carrying the same domains and particularly when the distribution of the domains is conserved, are expected to perform similar functions. Thus, similarities in chemical structure and in regulation (expression pattern, tissue specificity) could be useful to identify functional homologues of GRUBX.

"Homologues" of GRUBX encompass proteins having amino acid substitutions, insertions and/or deletions relative to the unmodified protein.

"Substitutional variants" of a protein are those in which at least one residue in an amino acid sequence has been removed and a different residue inserted in its place. Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; insertions will usually be of the order of about 1 to 10 amino acid residues, and deletions will range from about 1 to 20 residues. Preferably, amino acid substitutions comprise conservative amino acid substitutions.

"Insertional variants" of a protein are those in which one or more amino acid residues are introduced into a predetermined site in a protein. Insertions can comprise amino-terminal and/or carboxy-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than amino- or carboxy-terminal fusions, of the order of about 1 to 10 residues. Examples of amino- or carboxy-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, $(histidine)_6$-tag, glutathione S-transferase-tag, protein A, maltose-binding protein, dihydrofolate reductase, Tag·100 epitope, c-myc epitope, FLAG®-epitope, lacZ, CMP (calmodulin-binding peptide), HA epitope, protein C epitope and VSV epitope.

"Deletion variants" of a protein are characterised by the removal of one or more amino acids from the protein. Amino acid variants of a protein may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulations. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those skilled in the art and include M13 mutagenesis, T7-Gen *in vitro* mutagenesis (USB, Cleveland, OH), QuickChange Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols.

The term "derivatives" refers to peptides, oligopeptides, polypeptides, proteins and enzymes which may comprise substitutions, deletions or additions of naturally and non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of the protein, for example, as presented in SEQ ID NO: 2 or 4. "Derivatives" of GRUBX encompass peptides, oligopeptides, polypeptides, proteins and enzymes which may comprise naturally occurring altered, glycosylated, acylated or non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of the polypeptide. A derivative may also comprise one or more non-amino acid substituents compared to the amino acid sequence from which it is derived, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence such as, for example, a reporter molecule which is bound to facilitate its detection, and non-naturally occurring amino acid residues relative to the amino acid sequence of a naturally-occurring protein.

"Active fragments" of a GRUBX protein encompasses at least five contiguous amino acid residues of a protein, which residues retain similar biological and/or functional activity to the naturally occurring protein. The active fragment preferably also comprises a UBX and a PUG domain.

Advantageously, the method according to the present invention may also be practised using portions of a DNA or nucleic acid sequence, which portions retain GRUBX activity, i.e. a similar biological function to that of SEQ ID NO: 2. Portions of a DNA sequence refer to a piece of DNA derived or prepared from an original (larger) DNA molecule, which DNA portion, when expressed in a plant, gives rise to plants having modified growth characteristics. The portion may comprise many genes, with or without additional control elements, or may contain just spacer sequences etc.

The present invention also encompasses nucleic acid sequences capable of hybridising with a nucleic acid sequence encoding a GRUBX protein, which nucleic acid sequences may also be useful in practising the methods according to the invention. The term "hybridisation" as defined herein is a process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridisation process can occur entirely in solution, i.e. both complementary nucleic acids are in solution. Tools in molecular biology relying on such a process include the polymerase chain reaction (PCR; and all methods based thereon), subtractive hybridisation, random primer extension, nuclease S1 mapping, primer extension, reverse transcription, cDNA synthesis, differential display of RNAs, and DNA sequence determination. The hybridisation process can also occur with one of the complementary nucleic acids immobilised to a matrix such as magnetic beads, Sepharose beads or any other resin. Tools in molecular biology relying on such a process include the isolation of poly (A^+) mRNA. The hybridisation process can furthermore occur with one of the complementary nucleic acids immobilised to a solid support such as a nitro-cellulose or nylon membrane or immobilised by e.g. photolithography to, for example, a siliceous glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). Tools in molecular biology relying on such a process include RNA and DNA gel blot analysis, colony hybridisation, plaque hybridisation, *in situ* hybridisation and micro array hybridisation. In order to allow hybridisation to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration and hybridisation buffer composition.

For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C.

High stringency conditions for hybridisation thus include high temperature and/or low salt concentration (salts include NaCl and Na_3-citrate) but can also be influenced by the inclusion of formamide in the hybridisation buffer and/or lowering the concentration of compounds such as SDS (sodium dodecyl sulphate) in the hybridisation buffer and/or exclusion of compounds such as dextran sulphate or polyethylene glycol (promoting molecular crowding) from the hybridisation buffer. Sufficiently low stringency hybridisation conditions are particularly preferred for the isolation of nucleic acids homologous to the DNA sequences of the invention defined supra. Elements contributing to homology include allelism, degeneration of the genetic code and differences in preferred codon usage.

"Stringent hybridisation conditions" and "stringent hybridisation wash conditions" in the context of nucleic acid hybridisation experiments such as Southern and Northern hybridisations are sequence dependent and are different under different environmental parameters. For example, longer sequences hybridise specifically at higher temperatures. The T_m is the temperature under defined ionic strength and pH, at which 50% of the target sequence hybridises to a perfectly matched probe. Specificity is typically the function of post-hybridisation washes. Critical factors of such washes include the ionic strength and temperature of the final wash solution.

Generally, stringent conditions are selected to be about 50°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature under defined ionic strength and pH, at which 50% of the target sequence hybridises to a perfectly matched probe. The T_m is dependent upon the solution conditions and the base composition of the probe, and may be calculated using the following equation:

T_m= 79.8°C + (18.5xlog[Na^+]) + (58.4°Cx%[G+C]) – (820x(#bp in duplex)$^{-1}$) – (0.5x% formamide)

More preferred stringent conditions are when the temperature is 20°C below T_m, and the most preferred stringent conditions are when the temperature is 10°C below T_m. Non-specific binding may also be controlled using any one of a number of known techniques such as, for example, blocking the membrane with protein containing solutions, additions of heterologous RNA, DNA, and SDS to the hybridisation buffer, and treatment with Rnase.

Wash conditions are typically performed at or below hybridisation stringency. Generally, suitable stringent conditions for nucleic acid hybridisation assays or gene amplification detection procedures are as set forth above. More or less stringent conditions may also be selected.

For the purposes of defining the level of stringency, reference can conveniently be made to Sambrook et al. (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York or to Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989). An example of low stringency conditions is 4-6x SSC / 0.1-0.5% w/v SDS at 37-45°C for 2-3 hours. Depending on the source and concentration of the nucleic acid involved in the hybridisation, alternative conditions of stringency may be employed such as medium stringent conditions. Examples of medium stringent conditions include 1-4x SSC / 0.25% w/v SDS at ≥ 45°C for 2-3 hours. An example of high stringency conditions includes 0.1-1x SSC / 0.1% w/v SDS at 60°C for 1-3 hours. The skilled artisan is aware of various parameters which

may be altered during hybridisation and washing and which will either maintain or change the stringency conditions. For example, another stringent hybridisation condition is hybridisation at 4x SSC at 65°C, followed by a washing in 0.1x SSC, at 65°C for about one hour. Alternatively, an exemplary stringent hybridisation condition is in 50% formamide, 4x SSC at 42°C. Still another example of stringent conditions include hybridisation at 62°C in 6x SSC, 0.05x BLOTTO and washing at 2x SSC, 0.1% w/v SDS at 62°C.

The methods according to the present invention may also be practised using an alternative splice variant of a nucleic acid sequence encoding a GRUBX protein. The term "alternative splice variant" as used herein encompasses variants of a nucleic acid sequence in which selected introns and/or exons have been excised, replaced or added. Such variants will be ones in which the biological activity of the protein remains unaffected, which can be achieved by selectively retaining functional segments of the protein. Such splice variants may be found in nature or can be manmade. Methods for making such splice variants are well known in the art. Therefore according to another aspect of the present invention, there is provided, a method for modifying the growth characteristics of plants, comprising modulating expression in a plant of an alternative splice variant of a nucleic acid sequence encoding a GRUBX protein and/or by modulating activity and/or levels of a GRUBX protein encoded by the alternative splice variant. Preferably, the splice variant is a splice variant of the sequence represented by SEQ ID NO: 1.

Advantageously, the methods according to the present invention may also be practised using allelic variants of a nucleic acid sequence encoding a GRUBX protein, preferably an allelic variant of a sequence represented by SEQ ID NO: 1. Allelic variants exist in nature and encompassed within the methods of the present invention is the use of these natural alleles. Allelic variants encompass Single Nucleotide Polymorphisms (SNPs), as well as Small Insertion/Deletion Polymorphisms (INDELs). The size of INDELs is usually less than 100 bp). SNPs and INDELs form the largest set of sequence variants in naturally occurring polymorphic strains of most organisms.

The use of these allelic variants in particular conventional breeding programmes, such as in marker-assisted breeding is also encompassed by the present invention; this may be in addition to their use in the methods according to the present invention. Such breeding programmes sometimes require the introduction of allelic variations in the plants by mutagenic treatment of a plant. One suitable mutagenic method is EMS mutagenesis. Identification of allelic variants then may take place by, for example, PCR. This is followed by a selection step for selection of superior allelic variants of the GRUBX sequence in question and which give rise to altered growth characteristics in a plant. Selection is typically carried out by monitoring

growth performance of plants containing different allelic variants of the sequence in question, for example, different allelic variants of SEQ ID NO: 1. Monitoring growth performance can be done in a greenhouse or in the field. Further optional steps include crossing plants, in which the superior allelic variant was identified, with another plant. This could be used, for example, to make a combination of interesting phenotypic features.

Therefore, as mutations in the *GRUBX* gene may occur naturally, they may form the basis for selection of plants showing higher yield.

Accordingly, as another aspect of the invention, there is provided a method for the selection of plants having altered growth characteristics, which method is based on the selection of superior allelic variants of the GRUBX sequence and which give rise to altered growth characteristics in a plant.

The methods according to the present invention may also be practised by introducing into a plant at least a part of a (natural or artificial) chromosome (such as a Bacterial Artificial Chromosome (BAC)), which chromosome contains at least a gene/nucleic acid sequence encoding a GRUBX protein (such as SEQ ID NO: 1 or SEQ ID NO 3), preferably together with one or more related gene family members and/or nucleic acid sequence(s) encoding regulatory proteins for GRUBX expression and/or activity. Therefore, according to a further aspect of the present invention, there is provided a method for modifying the growth characteristics of plants by introducing into a plant at least a part of a chromosome comprising at least a gene/nucleic acid encoding a GRUBX protein.

According to another aspect of the present invention, advantage may be taken of the nucleotide sequence capable of modulating expression of a nucleic acid encoding a GRUBX protein in breeding programmes. The nucleic acid sequence may be on a chromosome, or a part thereof, comprising at least the nucleic acid sequence encoding the GRUBX protein and preferably also one or more related family members. In an example of such a breeding programme, a DNA marker is identified which may be genetically linked to a gene capable of modulating expression of a nucleic acid encoding a GRUBX protein in a plant, which gene may be a gene encoding the GRUBX protein itself or any other gene which may directly or indirectly influence expression of the gene encoding a GRUBX protein and/or activity of the GRUBX protein itself. This DNA marker may then be used in breeding programs to select plants having altered growth characteristics.

The present invention therefore extends to the use of a nucleic acid sequence in breeding programs, which nucleic acid sequence is capable of modulating expression of a nucleic acid

encoding a GRUBX protein or capable of modulating the activity and/or levels of a GRUBX protein.

Alternatively, the gene encoding a GRUBX protein itself can be used as a (genetic) marker to detect the presence or absence of a desired trait, or of a Quantitative Trait Locus (QTL). In this application of the invention, the gene encoding GRUBX is genetically linked to the desirable trait, and typically the phenotypes caused by the gene encoding a GRUBX are monitored in order to breed and select plants with that desired trait. This desirable trait or QTL, may comprise a single gene or a cluster of linked genes that affect the desirable trait.

It is standard practice in molecular biology to select upon transfection or transformation those individuals (or groups of individuals, such as bacterial or yeast colonies or phage plaques or eukaryotic cell clones) that are effectively transfected or transformed with the desired genetic construct. Typically these selection procedures are based on the presence of a selectable or screenable marker in the transfected genetic construct, to distinguish the positive individuals easily from the negative individuals. It is envisaged that the gene encoding a GRUBX protein can also be used for these purposes, since introduction of this gene into a host cell results in altered growth characteristics of said host cell.

Therefore, according to another feature of the present invention, there is provided a method for modifying plant growth characteristics, comprising modulating expression in a plant of a nucleic acid sequence encoding a GRUBX protein and/or modulating levels and/or activity of a GRUBX protein, wherein said nucleic acid sequence and said protein includes variants chosen from:

(i) an alternative splice variant of a nucleic acid sequence encoding an GRUBX protein or wherein said GRUBX protein is encoded by a splice variant;

(ii) an allelic variant of a nucleic acid sequence encoding an GRUBX protein or wherein said GRUBX protein is encoded by an allelic variant;

(iii) a nucleic acid sequence encoding a GRUBX protein and that is comprised on at least a part of an artificial chromosome, which artificial chromosome preferably also comprises one or more related gene family members;

(iv) a functional portion of a GRUBX encoding nucleic acid;

(v) sequence capable of hybridising to a GRUBX encoding nucleic acid;

(vi) homologues, derivatives and active fragments of a GRUBX protein.

According to a preferred aspect of the present invention, enhanced or increased expression of a nucleic acid is envisaged. Methods for obtaining enhanced or increased expression of

genes or gene products are well documented in the art and include, for example, overexpression driven by a (strong) promoter, the use of transcription enhancers or translation enhancers. The term overexpression as used herein means any form of expression that is additional to the original wild-type expression level. Preferably the nucleic acid to be introduced into the plant and/or the nucleic acid that is to be overexpressed in the plants is in a sense direction with respect to the promoter to which it is operably linked. Preferably, the nucleic acid to be overexpressed encodes a GRUBX protein, further preferably the nucleic acid sequence encoding the GRUBX protein is isolated from a dicotyledonous plant, preferably of the family Solanaceae, further preferably wherein the sequence is isolated from *Nicotiana tabacum*, most preferably the nucleic acid sequence is as represented by SEQ ID NO: 1 or a portion thereof, or encodes an amino acid sequence as represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof. Alternatively, the nucleic acid sequence encoding the GRUBX protein is as represented by MIPS No. At2g01650, SEQ ID NO: 3 or 6, or is a portion thereof, or encodes an amino acid sequence as represented by Q9ZU93, SEQ ID NO: 4 or 7, or encodes a homologue, derivative or active fragment thereof. It should be noted that the applicability of the invention does not rest upon the use of the nucleic acid represented by SEQ ID NO: 1, nor upon the nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 2, but that other nucleic acid sequences encoding homologues, derivatives or active fragments of SEQ ID NO: 2, or portions of SEQ ID NO: 1, or sequences hybridising with SEQ ID NO: 1 may be used in the methods of the present invention. In particular, the nucleic acids useful in the methods of the present invention encode proteins comprising at least a UBX and a PUG domain, and optionally also a Zinc finger domain.

Alternatively and/or additionally, increased expression of a GRUBX encoding gene or increased activities and/or levels of a GRUBX protein in a plant cells, is achieved by mutagenesis of the plant cell. For example these mutations can be responsible for the altered control of a GRUBX encoding gene, resulting in higher expression of the gene. Mutations can also cause conformational changes of the protein, resulting in higher activity and/or levels of the protein.

According to a further embodiment of the present invention, genetic constructs and vectors to facilitate introduction and/or expression of the nucleotide sequences useful in the methods according to the invention are provided. Therefore, according to a second embodiment of the present invention, there is provided a gene construct comprising:

(i) a nucleic acid sequence capable of increasing expression of a nucleic acid encoding a GRUBX protein and/or capable of increasing the activity and/or level of a GRUBX protein;

(ii) one or more control sequences capable of regulating expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

Constructs useful in the methods according to the present invention may be created using recombinant DNA technology well known to persons skilled in the art. The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming plants and suitable for expression of the gene of interest in the transformed cells.
The genetic construct can be an expression vector wherein the nucleic acid sequence is operably linked to one or more control sequences allowing expression in prokaryotic and/or eukaryotic host cells.

According to a preferred embodiment of the invention, the genetic construct is an expression vector designed to overexpress the nucleic acid sequence. The nucleic acid sequence capable of increasing expression of a nucleic acid encoding a GRUBX protein and/or activity of the GRUBX protein itself may be a nucleic acid sequence encoding a GRUBX protein or a homologue, derivative or active fragment thereof, such as any of the nucleic acid sequences described hereinbefore. A preferred nucleic acid sequence is the sequence represented by SEQ ID NO: 1 or a portion thereof or sequences capable of hybridising therewith or a nucleic acid sequence encoding a sequence represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof. Preferably, this nucleic acid is cloned in the sense orientation relative to the control sequence to which it is operably linked.

Plants are transformed with a vector comprising the sequence of interest (i.e., the nucleic acid sequence capable of modulating expression of nucleic acid encoding a GRUBX protein), which sequence is operably linked to one or more control sequences (at least a promoter). The terms "regulatory element", "control sequence" and "promoter" are all used herein interchangeably and are to be taken in a broad context to refer to regulatory nucleic acid sequences capable of effecting expression of the sequences to which they are ligated. Encompassed by the aforementioned terms are transcriptional regulatory sequences derived from a classical eukaryotic genomic gene (including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence) and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. Also included within the term is a transcriptional regulatory sequence of a classical prokaryotic gene, in which case it may include a –35 box sequence and/or –10 box transcriptional regulatory sequences. The term "regulatory element" also encompasses a

synthetic fusion molecule or derivative which confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ. The terms "control sequence", "regulatory sequence", "regulatory element" and "promoter" are used interchangeably herein. The term "operably linked" as used herein refers to a functional linkage between the promoter sequence and the gene of interest, such that the promoter sequence is able to initiate transcription of the gene of interest.

Advantageously, any type of promoter may be used to drive expression of the nucleic acid sequence depending on the desired outcome. Suitable promoters include promoters that are active in monocotyledonous plants such as rice or maize.

Preferably, the nucleic acid sequence capable of modulating expression of a gene encoding a GRUBX protein is operably linked to a seed-preferred promoter. The term "seed-preferred" as defined herein refers to a promoter that is expressed predominantly in seed tissue, but not necessarily exclusively in this tissue. The term "seed-preferred" encompasses all promoters that are active in seeds. Seed tissue encompasses any part of the seed including the endosperm, aleurone or embryo. Preferably, the seed-preferred promoter is a prolamine promoter, or a promoter of similar strength and/or a promoter with a similar expression pattern. Promoter strength and/or expression pattern can be analysed for example by coupling the promoter to a reporter gene and assay the expression of the reporter gene in various tissues of the plant. One suitable reporter gene well known to a person skilled in the art is bacterial *beta-glucuronidase.*
Examples of other seed-preferred promoters are presented in Table 1, and these promoters or derivatives thereof are useful for the methods of the present invention.

TABLE 1: Examples of seed-preferred promoters for use in the performance of the present invention:

GENE SOURCE	EXPRESSION PATTERN	REFERENCE
seed-specific genes	seed	Simon, *et al., Plant Mol. Biol. 5:* 191, 1985; Scofield, *et al., J. Biol. Chem. 262:* 12202, 1987.; Baszczynski, *et al., Plant Mol. Biol. 14:* 633, 1990.
Brazil Nut albumin	seed	Pearson, et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	seed	Ellis, et al., Plant Mol. Biol. 10: 203-214, 1988.

glutelin (rice)	seed	Takaiwa, et al., Mol. Gen. Genet. 208: 15-22, 1986; Takaiwa, et al., FEBS Letts. 221: 43-47, 1987.
zein	seed	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	seed	Stalberg, *et al, Planta 199:* 515-519, 1996.
wheat LMW and HMW glutenin-1	endosperm	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	seed	Albani *et al,* Plant Cell, 9: 171-184, 1997
wheat α, β, γ-gliadins	endosperm	EMBO J. 3:1409-15, 1984
barley *ltr1* promoter	endosperm	
barley B1, C, D, hordein	endosperm	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	endosperm	Mena *et al,* The Plant Journal, 116(1): 53-62, 1998
blz2	endosperm	EP99106056.7
synthetic promoter	endosperm	Vicente-Carbajosa *et al., Plant J. 13:* 629-640, 1998.
rice prolamin NRP33	endosperm	Wu *et al,* Plant Cell Physiology 39(8) 885-889, 1998
rice α-globulin Glb-1	endosperm	Wu *et al,* Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	embryo	Sato *et al,* Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
rice α-globulin REB/OHP-1	endosperm	Nakase *et al.* Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose PP	endosperm	Trans Res 6:157-68, 1997
maize ESR gene family	endosperm	Plant J 12:235-46, 1997
sorgum γ-kafirin	endosperm	PMB 32:1029-35, 1996
KNOX	embryo	Postma-Haarsma *et al,* Plant Mol. Biol. 39:257-71, 1999
rice oleosin	embryo and aleuron	Wu *et at,* J. Biochem., 123:386, 1998
sunflower oleosin	seed (embryo and dry seed)	Cummins, *et al.*, Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	weak in endosperm	
PRO0135, rice alpha-globulin	strong in endosperm	

PRO0136, rice alanine aminotransferase	weak in endosperm	
PRO0147, trypsin inhibitor ITR1 (barley)	weak in endosperm	
PRO0151, rice WSI18	embryo + stress	
PRO0175, rice RAB21	embryo + stress	
PRO0218, rice oleosin 18kd	aleurone + embryo	

Optionally, one or more terminator sequences may also be used in the construct introduced into a plant. The term "terminator" encompasses a control sequence which is a DNA sequence at the end of a transcriptional unit which signals 3' processing and polyadenylation of a primary transcript and termination of transcription. Additional regulatory elements may include transcriptional as well as translational enhancers. Those skilled in the art will be aware of terminator and enhancer sequences which may be suitable for use in performing the invention. Such sequences would be known or may readily be obtained by a person skilled in the art.

The genetic constructs of the invention may further include an origin of replication sequence which is required for maintenance and/or replication in a specific cell type. One example is when a genetic construct is required to be maintained in a bacterial cell as an episomal genetic element (e.g. plasmid or cosmid molecule). Preferred origins of replication include, but are not limited to, the f1-ori and colE1.

The genetic construct may optionally comprise a selectable marker gene. As used herein, the term "selectable marker gene" includes any gene which confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells which are transfected or transformed with a nucleic acid construct of the invention. Suitable markers may be selected from markers that confer antibiotic or herbicide resistance, that introduce a new metabolic trait or that allow visual selection. Examples of selectable marker genes include genes conferring resistance to antibiotics (such as *npt*II that phosphorylates neomycin and kanamycin, or *hpt*, phosphorylating hygromycin), to herbicides (for example *bar* which provides resistance to Basta; *aroA* or *gox* providing resistance against glyphosate), or genes that provide a metabolic trait (such as *manA*, allowing plants to use mannose as sole carbon source). Visual marker genes result in the formation of colour (for example β-glucuronidase, GUS), luminescence (such as luciferase) or fluorescence (Green Fluorescent Protein, GFP, and derivatives thereof).

In a preferred embodiment, the genetic construct as mentioned above, comprises a GRUBX in sense orientation coupled to a promoter that is preferably a seed-preferred promoter, such as for example the rice prolamine promoter. Therefore, another aspect of the present invention is a vector construct carrying an expression cassette essentially similar to SEQ ID NO 5, comprising a prolamine promoter, the *Nicotiana tabacum GRUBX* gene and the T-zein + T-rubisco deltaGA transcription terminator sequence. A sequence essentially similar to SEQ ID NO 5 encompasses a first nucleic acid sequence encoding a protein homologous to SEQ ID NO 2 or hybridising to SEQ ID NO 1, which first nucleic acid is operably linked to a prolamine promoter or a promoter with a similar expression pattern, additionally or alternatively the first nucleic acid is linked to a transcription termination sequence.

Therefore according to another aspect of the invention, there is provided a nucleic acid construct, comprising an expression cassette in which is located a nucleic acid sequence encoding an GRUBX protein, chosen from the group comprising:

(i) a nucleic acid sequence represented by SEQ ID NO: 1 or the complement strand thereof;

(ii) a nucleic acid sequence encoding an amino acid sequence represented by SEQ ID NO: 2 or homologues, derivatives or active fragments thereof;

(iii) a nucleic acid sequence capable of hybridising (preferably under stringent conditions) with a nucleic acid sequence of (i) or (ii) above, which hybridising sequence preferably encodes a protein having GRUBX protein activity;

(iv) a nucleic acid sequence according to (i) to (iii) above which is degenerate as a results of the genetic code;

(v) nucleic acid sequence which is an allelic variant to the nucleic acid sequences according to (i) to (iv);

(vi) nucleic acid sequence which is an alternative splice variant to the nucleic acid sequences according to (i) to (v);

The present invention also encompasses plants obtainable by the methods according to the present invention. The present invention therefore provides plants obtainable by the method according to the present invention, which plants have modified growth characteristics and which plants have altered GRUBX protein activity and/or levels and/or altered expression of a nucleic acid encoding a GRUBX protein.

According to a third embodiment of the present invention, there is provided a method for the production of transgenic plants having modified growth characteristics, comprising introduction and expression in a plant of a nucleic acid molecule of the invention.

More specifically, the present invention provides a method for the production of transgenic plants having modified growth characteristics, which method comprises:

(i) introducing into a plant cell a nucleic acid sequence capable of modulating expression of a gene encoding a GRUBX protein and/or capable of modulating the activity and/or levels of a GRUBX protein;

(ii) cultivating the plant cell under conditions promoting plant growth.

The GRUBX protein itself and/or the *GRUBX* nucleic acid itself may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of the plant). According to a preferred feature of the present invention, the nucleic acid is preferably introduced into a plant by transformation. The nucleic acid is preferably as represented by SEQ ID NO: 1 or a portion thereof or sequences capable of hybridising therewith, or is a nucleic acid encoding an amino acid sequence represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof. Alternatively, the nucleic acid sequence is as represented by any of MIPS No. At2g01650, SEQ ID NO: 3, SEQ ID NO 6, or by a portion thereof or by sequences capable of hybridising with any of the aforementioned sequences. The amino acid sequence may alternatively be a sequence as represented by any of SPTrEMBL Q9ZU93, GenBank Acc. Nr. AAR01744, SEQ ID NO: 4, SEQ ID NO 7, or by homologues, derivatives or active fragments thereof.

The term "transformation" as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell can then be used to regenerate a transformed plant in a manner known to persons skilled in the art.

Transformation of a plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable

ancestor cell. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. et al., 1882, Nature 296, 72-74; Negrutiu I. et al., June 1987, Plant Mol. Biol. 8, 363-373); electroporation of protoplasts (Shillito R.D. et al., 1985 Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A. et al., 1986, Mol. Gen Genet 202, 179-185); DNA or RNA-coated particle bombardment (Klein T.M. et al., 1987, Nature 327, 70) infection with (non-integrative) viruses and the like. Transgenic rice plants expressing an *GRUBX* gene are preferably produced via *Agrobacterium*-mediated transformation using any of the well known methods for rice transformation, such as described in any of the following: published European patent application EP 1198985 A1, Aldemita and Hodges (Planta, 199, 612-617, 1996); Chan et al. (Plant Mol. Biol. 22 (3) 491-506, 1993), Hiei et al. (Plant J. 6 (2) 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida et al. (Nat. Biotechnol. 1996 Jun; 14(6): 745-50) or Frame et al. (Plant Physiol. 2002 May; 129(1): 13-22), which disclosures are incorporated by reference herein as if fully set forth.

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant.

Following DNA transfer and regeneration, putatively transformed plants may be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed to give homozygous second generation (or T2) transformants, and the T2 plants further propagated through classical breeding techniques.

The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

The present invention clearly extends to any plant cell or plant produced by any of the methods described herein, and to all plant parts, propagules and progeny thereof. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced in the parent by the methods according to the invention. The invention also includes host cells containing an isolated nucleic acid molecule encoding a protein capable of modulating levels and/or activity of GRUBX, preferably wherein the protein is a GRUBX protein. Preferred host cells according to the invention are plant cells. Therefore, the invention also encompasses host cells or transgenic plants having altered growth characteristics, characterized in that said host cell or transgenic plant has modulated expression of a nucleic acid sequence encoding a GRUBX protein and/or modulated activity and/or level of a GRUBX protein.

The invention also extends to harvestable parts of a plant such as but not limited to seeds, leaves, fruits, flowers, stems or stem cultures, rhizomes, roots, tubers and bulbs.

The term "plant" as used herein encompasses whole plants, ancestors and progeny of the plants, plant parts, plant cells, tissues and organs. The term "plant" also therefore encompasses suspension cultures, embryos, meristematic regions, callus tissue, leaves, flowers, fruits, seeds, roots (including rhizomes and tubers), shoots, bulbs, stems, gametophytes, sporophytes, pollen, and microspores. Plants that are particularly useful in the methods of the invention include algae, ferns, and all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants, including fodder or forage legumes, ornamental plants, food crops, trees, or shrubs selected from the list comprising *Abelmoschus* spp., *Acer* spp., *Actinidia* spp., *Agropyron* spp., *Allium* spp., *Amaranthus* spp., *Ananas comosus, Annona* spp., *Apium graveolens, Arabidopsis thaliana, Arachis* spp, *Artocarpus* spp., *Asparagus officinalis, Avena sativa, Averrhoa carambola, Benincasa hispida, Bertholletia excelsea, Beta vulgaris, Brassica* spp., *Cadaba farinosa, Camellia sinensis, Canna indica, Capsicum* spp., *Carica papaya, Carissa macrocarpa, Carthamus tinctorius, Carya* spp., *Castanea* spp., *Cichorium endivia, Cinnamomum spp., Citrullus lanatus, Citrus* spp., *Cocos* spp., *Coffea* spp., *Cola* spp., *Colocasia esculenta, Corylus* spp., *Crataegus* spp., *Cucumis* spp., *Cucurbita* spp., *Cynara* spp., *Daucus carota, Desmodium* spp., *Dimocarpus longan, Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Eleusine coracana, Eriobotrya japonica, Eugenia uniflora, Fagopyrum* spp., *Fagus* spp., *Ficus carica, Fortunella spp., Fragaria* spp., *Ginkgo biloba, Glycine* spp., *Gossypium hirsutum, Helianthus*

spp., *Hibiscus* spp., *Hordeum* spp., *Ipomoea batatas, Juglans* spp., *Lactuca sativa, Lathyrus* spp., *Lemna spp., Lens culinaris, Linum usitatissimum, Litchi chinensis, Lotus* spp., *Luffa acutangula, Lupinus* spp., *Macrotyloma* spp., *Malpighia emarginata, Malus* spp., *Mammea americana, Mangifera indica, Manihot* spp., *Manilkara zapota, Medicago sativa, Melilotus* spp., *Mentha* spp., *Momordica* spp., *Morus nigra, Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Ornithopus* spp., *Oryza* spp., *Panicum miliaceum, Passiflora edulis, Pastinaca sativa, Persea* spp., *Petroselinum crispum, Phaseolus* spp., *Phoenix* spp., *Physalis* spp., *Pinus* spp., *Pistacia vera, Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus* spp., *Psidium* spp., *Punica granatum, Pyrus communis, Quercus* spp., *Raphanus sativus, Rheum rhabarbarum, Ribes* spp., *Rubus* spp., *Saccharum* spp., *Sambucus* spp., *Secale cereale, Sesamum* spp., *Solanum* spp., *Sorghum bicolor, Spinacia* spp., *Syzygium* spp., *Tamarindus indica, Theobroma cacao, Trifolium* spp., *Triticosecale rimpaui, Triticum* spp., *Vaccinium* spp., *Vicia* spp., *Vigna spp., Vitis* spp., *Zea mays, Zizania palustris, Ziziphus* spp., amongst others.

According to a preferred feature of the present invention, the plant is a crop plant comprising soybean, sunflower, canola, alfalfa, rapeseed or cotton. Further preferably, the plant according to the present invention is a monocotyledonous plant such as sugarcane, most preferably a cereal, such as rice, maize, wheat, millet, barley.

However, it is envisaged that the methods of the present invention can be applied to a wide variety of plants, since the domain conservation among the known eukaryotic GRUBX homologues suggests an equally conserved function in cellular metabolism.

Advantageously, performance of the methods according to the present invention results in plants having a variety of modified growth characteristics, such modified growth characteristics including modified growth, modified yield/biomass, modified architecture and a modified cell division, each relative to corresponding wild type plants. Preferably, the modified growth characteristics are improved growth characteristics and include increased yield/biomass, and improved architecture, each relative to corresponding wild type plants.

The present invention relates to methods to alter growth characteristics of a plant or to methods to produce plants with altered growth characteristics, wherein the growth characteristics comprise any one or more selected from: increased yield, increased biomass, increased total above ground area, increased plant height, increased number of tillers, increased number of first panicles, increased number of second panicles, increased total number of seeds, increased number of filled seeds, increased total seed yield per plant, increased harvest index, increased thousand kernel weight, increased Tmid, increased T45 or A90, increased A42, altered cycling time and/or an altered growth curve. The present invention also provides methods to alter one of the above mentioned growth characteristics, without

causing a penalty on one of the other growth characteristics, for example increase of the above ground green tissue area while retaining the same number of filled seeds and the same seed yield.

The term "increased yield" encompasses an increase in biomass in one or more parts of a plant relative to the biomass of corresponding wild-type plants. The term also encompasses an increase in seed yield, which includes an increase in the biomass of the seed (seed weight) and/or an increase in the number of (filled) seeds and/or in the size of the seeds and/or an increase in seed volume, each relative to corresponding wild-type plants. For maize, the increase of seed yield may be reflected in an increase of rows (of seeds) per ear and/or an increased number of kernels per row. An increase in seed size and/or volume may also influence the composition of seeds. An increase in seed yield could be due to an increase in the number and/or size of flowers. An increase in yield might also increase the harvest index, which is expressed as a ratio of the total biomass over the yield of harvestable parts, such as seeds; or Thousand Kernel Weight. Increased yield also encompasses the capacity for planting at higher density (number of plants per hectare or acre).

The term "modified cell division" encompasses an increase or decrease in cell division or an abnormal cell division/cytokinesis, altered plane of division, altered cell polarity, altered cell differentiation. The term also comprises phenomena such as endomitosis, acytokinesis, polyploidy, polyteny and endoreduplication.

It can be envisaged that plants having increased biomass and height exhibit a modified growth rate when compared to corresponding wild-type plants. The term "modified growth rate" as used herein encompasses, but is not limited to, a faster rate of growth in one or more parts of a plant (including green biomass and including seeds), at one or more stages in the life cycle of a plant. The term "modified growth" encompasses enhanced vigour, earlier flowering, modified cycling time. Plants with modified growth may show a modified growth curve and may have modified values for their T_{mid} or T_{90} (respectively the time needed to reach half of their maximal area or 90% of their area, each relative to corresponding wild-type plants).

According to a preferred feature of the present invention, performance of the methods according to the present invention result in plants having modified yield. Preferably, the modified yield includes at least an increase in harvest index, relative to control plants. Therefore, according to the present invention, there is provided a method for increasing yield of plants, in particular harvest index, which method comprises modulating expression of a nucleic acid sequence encoding a GRUBX protein and/or modulating activity of a GRUBX

protein itself in a plant, preferably wherein the GRUBX protein is encoded by a nucleic acid sequence represented by SEQ ID NO: 1 or a portion thereof or by sequences capable of hybridising therewith or wherein the GRUBX protein is represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof. Alternatively, the GRUBX may be encoded by a nucleic acid sequence represented by any of MIPS No. At2g01650, SEQ ID NO: 3, or by a portion thereof or by sequences capable of hybridising therewith, or wherein the GRUBX is represented by any of SPTrEMBL Q9ZU93, SEQ ID NO: 4, or a homologue, derivative or active fragment of any thereof.

The methods of the present invention are favourable to apply to crop plants because the methods of the present invention are used to increase the harvest index of a plant. Therefore, the methods of the present invention are particularly useful for crop plants cultivated for their seeds, such as cereals. Accordingly, a particular embodiment of the present invention relates to a method to increase the harvest index of a cereal.

"Modified architecture" may be due to change in cell division. The term "architecture" as used herein encompasses the appearance or morphology of a plant, including any one or more structural features or combination of structural features thereof. Such structural features include the shape, size, number, position, texture, arrangement, and pattern of any cell, tissue or organ or groups of cells, tissues or organs of a plant, including the root, leaf, shoot, stem or tiller, petiole, trichome, flower, inflorescence (for monocotyledonous and dicotyledonous plants), panicles, petal, stigma, style, stamen, pollen, ovule, seed, embryo, endosperm, seed coat, aleurone, fibre, cambium, wood, heartwood, parenchyma, aerenchyma, sieve elements, phloem or vascular tissue, amongst others. Modified architecture therefore includes all aspects of modified growth of the plant.

The present invention also relates to the use of a nucleic acid encoding a GRUBX protein and to the use of portions thereof or nucleic acids hybridising therewith in modifying the growth characteristics of plants, preferably in increasing the yield and/or biomass of a plant. The present invention also relates to use of a GRUBX protein and to the use of homologues, derivatives and active fragments thereof in modifying the growth characteristics of plants. The nucleic acid sequence is preferably as represented by SEQ ID NO: 1 or a portion thereof or sequences capable of hybridising therewith or encodes an amino acid sequence represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof.

The present invention also relates to the use of a nucleic acid sequence encoding a GRUBX protein and to the use of portions thereof or nucleic acids hybridising therewith and to the use

of the GRUBX protein itself and of homologues, derivatives and active fragments thereof as growth regulators. The nucleic acid sequences hereinbefore described (and portions of the same and sequences capable of hybridising with the same) and the amino acid sequences hereinbefore described (and homologues, derivatives and active fragments of the same) are useful in modifying the growth characteristics of plants, as hereinbefore described. The sequences would therefore find use as growth regulators, to stimulate or inhibit plant growth. Therefore, the present invention provides a composition comprising a GRUBX protein or a protein represented by SEQ ID NO 2 or a homologue, derivative or active fragment thereof for use in modifying the growth characteristics of plants. The present invention furthermore provides a composition comprising a nucleic acid encoding a GRUBX protein, or a nucleic acid as represented by SEQ ID NO 1 or a portion thereof or a sequence hybridising therewith for use in modifying the growth characteristics of plants. The present invention also provides a composition comprising a protein represented by any of the aforementioned amino acid sequences or homologues, derivatives or active fragments thereof for the use as a growth regulator.

Conversely, the sequences according to the present invention may also be interesting targets for agrochemical compounds, such as herbicides or growth stimulators. Accordingly, the present invention encompasses use of the aforementioned nucleic acid sequences (or a portion of the same or sequences capable of hybridising with the same) or an amino acid sequence as hereinbefore described (or homologues, derivatives and active fragments of the same) as targets for an agrochemical compound, such as an herbicide or a growth stimulator.

The contents of all references included in this application are incorporated by reference herein as if fully set forth.

The present invention will now be described with reference to the following figures in which:

Figure 1a. Phylogenetic tree representing *Arabidopsis thaliana* proteins and animal reference proteins comprising a UBX domain, as recognised by the SMART tool. The human proteins are represented by their GenBank Accession numbers NP_079517 (*Homo sapiens* UBX domain containing 1 (UBXD1)), AAP97263 (*Homo sapiens* Fas-associated protein factor FAF1 mRNA), NP_005662 (*Homo sapiens* reproduction 8 (D8S2298E), REP8) and a rat protein by NP_114187 (*Rattus norvegicus* p47 protein). The other identifiers (except for SEQ ID NO 2, SEQ ID NO 4 and SEQ ID NO 7) are GenBank or SPTrEMBL accession numbers for *Arabidopsis thaliana* proteins.

Figur 1b. Phylogenetic tree representing plant proteins comprising a PUG domain, as recognised by the SMART tool. SEQ ID NO 2 and SEQ ID NO 4 are compared with

Arabidopsis thaliana proteins (SPTrEMBL accessions Q9ZU93 (Expressed protein), Q9FKI1 (Similarity to zinc metalloproteinase), Q9MAT3 (F13M7.16 protein), Q9FKC7 (Genomic DNA, chromosome 5, TAC clone:K24G6), Q9SF12 (Hypothetical protein), Q9C5S2 (Endoribonuclease/protein kinase IRE1), Q8RX75 (AT5g24360/K16H17_7), Q94IG5 (Ire1 homolog-1)), and with the rice protein SPTrEMBL Q7XIT1 (OsIre1p).

Figure 2a. Definition of UBX1 and PUG domains by their consensus sequences (SMART database). CONSENSUS/50%, respectively /65% and /80% are the consensus sequences for the top 50, 65 and 80% of the reference sequences comprising the UBX1 or PUG domain. The capital letters are the standard single letter IUPAC codes for the various amino acids, the other letters symbolise the nature of the amino acids as outlined below:

Class	Key	Residues
Alcohol	o	S,T
Aliphatic	l	I,L,V
Any	.	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
Aromatic	a	F,H,W,Y
Charged	c	D,E,H,K,R
Hydrophobic	h	A,C,F,G,H,I,K,L,M,R,T,V,W,Y
Negative	-	D,E
Polar	p	C,D,E,H,K,N,Q,R,S,T
Positive	+	H,K,R
Small	s	A,C,D,G,N,P,S,T,V
Tiny	u	A,G,S
Turnlike	t	A,C,D,E,G,H,K,N,Q,R,S,T

Figure 2b. UBX and PUG domain sequences present in SEQ ID NO 2 and in Q9ZU93.

Figure 2c. Alignment of Q9ZU93 and SEQ ID NO 2, PUG domains underlined, UBX domains in bold.

Figure 2d. Alignment of SEQ ID NO 2 and SEQ ID NO 4, PUG domains underlined, UBX domains in bold.

Figure 2e. Alignment of SEQ ID NO 4 and SEQ ID NO 7, PUG domains underlined, UBX domains in bold.

Figure 3. Schematic presentation of the entry clone p77, containing CDS0669 within the AttL1 and AttL2 sites for Gateway® cloning in the pDONR201 backbone. CDS0669 is the internal code for the tobacco *GRUBX* coding sequence. This vector contains also a bacterial kanamycine-resistance cassette and a bacterial origin of replication.

Figur 4. Binary vector for the expression in *Oryza sativa* of the tobacco GRUBX gene (CDS0669) under the control of the prolamine promoter (PRO0090). This vector contains a T-DNA derived from the Ti Plasmid, limited by a left border (LB repeat, LB Ti C58) and a right border (RB repeat, RB Ti C58)). From the left border to the right border, this T-DNA contains: a cassette for antibiotic selection of transformed plants; a cassette for visual screening of transformed plants; the PRO0090 - CDS0669 -zein and rbcS-deltaGA double terminator cassette for expression of the tobacco *GRUBX* gene. This vector also contains an origin of replication from pBR322 for bacterial replication and a selectable marker (Spe/SmeR) for bacterial selection with spectinomycin and streptomycin.

Figure 5. Examples of sequences useful in the present invention.

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration alone.

DNA manipulation: unless otherwise stated, recombinant DNA techniques are performed according to standard protocols described in (Sambrook (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York) or in Volumes 1 and 2 of Ausubel et al. (Current Protocols in Molecular Biology. New York: John Wiley and Sons, 1998). Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfase (1993) by R.D.D. Croy, published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

Example 1: Cloning of the CDS0669 sequence

Cloning of the *GRUBX* gene fragment from tobacco

A cDNA-AFLP experiment was performed on a synchronized tobacco BY2 cell culture (*Nicotiana tabacum* L. cv. Bright Yellow-2), and BY2 expressed sequence tags that were cell cycle modulated were elected for further cloning. The expressed sequence tags were used to screen a tobacco cDNA library and to isolate the full-length cDNA of interest, namely one coding for *GRUBX* gene (CDS0669).

Synchronization of BY2 cells.

A tobacco BY2 (*Nicotiana tabacum* L. cv. Bright Yellow-2) cultured cell suspension was synchronized by blocking cells in early S-phase with aphidicolin as follows. The cell suspension of *Nicotiana tabacum* L. cv. Bright Yellow 2 was maintained as described (Nagata et al. Int. Rev. Cytol. 132, 1-30, 1992). For synchronization, a 7-day-old stationary culture was diluted 10-fold in fresh medium supplemented with aphidicolin (Sigma-Aldrich, St. Louis, MO; 5 mg/l), a DNA-polymerase α inhibiting drug. After 24 h, cells were released from the block by several washings with fresh medium whereafter their cell cycle progression resumed.

RNA extraction and cDNA synthesis.

Total RNA was prepared by using LiCl precipitation (Sambrook et al, 2001) and poly(A^+) RNA was extracted from 500 µg of total RNA using Oligotex columns (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Starting from 1 µg of poly(A^+) RNA, first-strand cDNA was synthesized by reverse transcription with a biotinylated oligo-dT_{25} primer (Genset, Paris, France) and Superscript II (Life Technologies, Gaithersburg, MD). Second-strand synthesis was done by strand displacement with *Escherichia coli* ligase (Life Technologies), DNA polymerase I (USB, Cleveland, OH) and RNAse-H (USB).

cDNA-AFLP analysis.

Five hundred ng double-stranded cDNA was used for AFLP analysis as described (Vos *et al.*, Nucleic Acids Res. 23 (21) 4407-4414, 1995; Bachem et al., Plant J. 9 (5) 745-53, 1996) with modifications. The restriction enzymes used were *Bsf*YI and *Mse*I (Biolabs) and the digestion was done in two separate steps. After the first restriction digest with one of the enzymes, the 3' end fragments were trapped on Dyna beads (Dynal, Oslo, Norway) by means of their biotinylated tail, while the other fragments were washed away. After digestion with the second enzyme, the released restriction fragments were collected and used as templates in the subsequent AFLP steps. For pre-amplifications, a *Mse*I primer without selective nucleotides was combined with a *Bsf*YI primer containing either a T or a C as 3' most nucleotide. PCR conditions were as described (Vos et al., 1995). The obtained amplification mixtures were diluted 600-fold and 5 µl was used for selective amplifications using a P^{33}-labeled *Bsf*YI primer and the Amplitaq-Gold polymerase (Roche Diagnostics, Brussels, Belgium). Amplification products were separated on 5% polyacrylamide gels using the Sequigel system (Biorad). Dried gels were exposed to Kodak Biomax films as well as scanned in a PhosphorImager (Amersham Pharmacia Biotech, Little Chalfont, UK).

Characterization of AFLP fragments.

Bands corresponding to differentially expressed transcripts, among which the (partial) transcript corresponding to SEQ ID NO 1 (or CDS0669), were isolated from the gel and eluted DNA was reamplified under the same conditions as for selective amplification. Sequence information was obtained either by direct sequencing of the reamplified polymerase chain reaction product with the selective *Bsf*YI primer or after cloning the fragments in pGEM-T easy (Promega, Madison, WI) and sequencing of individual clones. The obtained sequences were compared against nucleotide and protein sequences present in the publicly available databases by BLAST sequence alignments (Altschul *et al.*, Nucleic Acids Res. 25 (17) 3389-3402 1997). When available, tag sequences were replaced with longer EST or isolated cDNA sequences to increase the chance of finding significant homology. The physical cDNA clone corresponding to SEQ ID NO 1 (CDS0669) was subsequently amplified from a commercial tobacco cDNA library as follows:

Cloning of the *GRUBX* gene (CDS0669)

A c-DNA library with an average size of inserts of 1,400 bp was prepared from poly(A^+) RNA isolated from actively dividing, non-synchronized BY2 tobacco cells. These library-inserts were cloned in the vector pCMVSPORT6.0, comprising a attB Gateway cassette (Life Technologies). From this library 46,000 clones were selected, arrayed in 384-well microtiter plates, and subsequently spotted in duplicate on nylon filters. The arrayed clones were screened by using pools of several hundreds of radioactively labelled tags as probe (including the BY2-tag corresponding to the sequence CDS0669, SEQ IDNO 1). Positive clones were isolated (among which the clone corresponding to CDS0669, SEQ I NO 1), sequenced, and aligned with the tag sequence. Alternatively, when the hybridisation with the tag would fail, the full-length cDNA corresponding to the tag was selected by PCR amplification: tag-specific primers were designed using primer3 program (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) and used in combination with a common vector primer to amplify partial cDNA inserts. Pools of DNA from 50.000, 100.000, 150.000, and 300.000 cDNA clones were used as templates in the PCR amplifications. Amplification products were then isolated from agarose gels, cloned, sequenced and their sequence aligned with those of the tags.

Next, the full-length cDNA corresponding to the nucleotide sequence of SEQ ID NO 1 was cloned from the pCMVsport6.0 library vector into pDONR201, a Gateway® donor vector (Invitrogen, Paisley, UK) via a LR reaction, resulting in the entry clone p77 (Figure 3).

Example 2: V ctor construction

The entry clone p77 was subsequently used in an LR reaction with p0830, a destination vector used for *Oryza sativa* transformation. This vector contains as functional elements within the T-DNA borders: a plant selectable marker; a visual marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the sequence of interest already cloned in the entry clone. A prolamine promoter for seed-preferred expression (PRO0090) is located upstream of this Gateway cassette. After the LR recombination step, the resulting expression vector p72 (Figure 4) can be transformed into the *Agrobacterium* strain LBA4404 and subsequently to *Oryza sativa* plants.

Example 3: Transformation of rice with the PRO0090-CDS0669 construct

Mature dry seeds of *Oryza sativa* japonica cultivar Nipponbare were dehusked. Sterilization was done by incubating the seeds for one minute in 70% ethanol, followed by 30 minutes in 0.2% $HgCl_2$ and by 6 washes of 15 minutes with sterile distilled water. The sterile seeds were then germinated on a medium containing 2,4-D (callus induction medium). After a 4-week incubation in the dark, embryogenic, scutellum-derived calli were excised and propagated on the same medium. Two weeks later, the calli were multiplied or propagated by subculture on the same medium for another 2 weeks. 3 days before co-cultivation, embryogenic callus pieces were sub-cultured on fresh medium to boost cell division activity. The *Agrobacterium* strain LBA4414 harbouring the binary vector p72 was used for co-cultivation. The *Agrobacterium* strain was cultured for 3 days at 28°C on AB medium with the appropriate antibiotics. The bacteria were then collected and suspended in liquid co-cultivation medium at an OD_{600} of about 1. The suspension was transferred to a petri dish and the calli were immersed in the suspension during 15 minutes. Next, the callus tissues were blotted dry on a filter paper, transferred to solidified co-cultivation medium and incubated for 3 days in the dark at 25°C. Thereafter, co-cultivated callus was grown on 2,4-D-containing medium for 4 weeks in the dark at 28°C in the presence of a selective agent at a suitable concentration. During this period, rapidly growing resistant callus islands developed. Upon transfer of this material to a regeneration medium and incubation in the light, the embryogenic potential was released and shoots developed in the next four to five weeks. Shoots were excised from the callus and incubated for 2 to 3 weeks on an auxin-containing medium from which they were transferred to soil. Hardened shoots were grown under high humidity and short days in a greenhouse. Finally seeds were harvested three to five months after transplanting. The method yielded single locus transformants at a rate of over 50 % (Aldemita and Hodges, Planta 199, 612-617, 1996; Chan et al., Plant Mol. Biol. 22(3), 491-506, 1993; Hiei et al., Plant J. 6(2), 271-282, 1994).

Exampl 4: Evaluation of transgenic ric transform d with th PRO0090-CDS0669 construct

Approximately 15 to 20 independent T0 rice transformants were generated. The primary transformants were transferred from tissue culture chambers to a greenhouse for growing and harvest of T1 seed. 6 events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes), and approximately 10 T1 seedlings lacking the transgene (nullizygotes), were selected by monitoring visual marker expression. A number of parameters related to vegetative growth and seed production were evaluated and all data were statistically analysed as outlined below:

Statistical analysis: t-test and F-test:

A two factor ANOVA (analysis of variants) was used as statistical model for the overall evaluation of plant phenotypic characteristics. An F-test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F-test is carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also named herein "global gene effect". If the value of the F-test shows that the data are significant, than it is concluded that there is a "gene" effect, meaning that not only presence or the position of the gene is causing the differences in phenotype. The threshold for significance for a true global gene effect is set at 5% probability level for the F-test.

To check for an effect of the genes within an event, i.e., for a line-specific effect, a t-test was performed within each event using data sets from the transgenic plants and the corresponding null plants. "Null plants" or "Null segregants" are the plants treated in the same way as the transgenic plant, but from which the transgene has segregated. Null plants can also be described as the homozygous negative transformants. The threshold for significance for the t-test is set at 10% probability level. Within one population of 4 transformation events, some events can be under or above this t-test threshold. This is based on the hypothesis that a gene might only have an effect in certain positions in the genome, and that the occurrence of this position-dependent effect is not uncommon. This kind of gene effect is also named herein a "line effect of the gene". The p-value is obtained by comparing the t-value to the t-distribution or alternatively, by comparing the F-value to the F-distribution. The p-value then stands for the probability that the null hypothesis (null hypothesis being "there is no effect of the transgene") is correct.

4.1 Vegetative growth measurements:

The selected T1 plants (approximately 10 with the transgene and approximately 10 without the transgene) were transferred to a greenhouse. Each plant received a unique barcode label to link unambiguously the phenotyping data to the corresponding plant. The selected T1 plants were grown on soil in 10 cm diameter pots under the following environmental settings: photoperiod= 11.5 h, daylight intensity= 30,000 lux or more, daytime temperature= 28°C or higher, night time temperature= 22°C, relative humidity= 60-70%. Transgenic plants and the corresponding nullizygotes were grown side-by-side at random positions. From the stage of sowing until the stage of maturity each plant was passed several times through a digital imaging cabinet and imaged. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles. Several parameters can be derived in an automated way from all the digital images of all the plants, using image analysis software.

4.2 Seed-related parameter measurements:

The mature primary panicles were harvested, bagged, barcode-labelled and then dried for three days in the oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. This procedure allows to derive a set of seed-related parameters.

Harvest index of plants

The harvest index in the present invention is defined as the ratio between the total seed yield and the above ground area (mm^2), multiplied by a factor 10^6. The total seed yield per plant was measured by weighing all filled husks harvested from a plant as described above. Plant aboveground area was determined by counting the total number of pixels of the digital images from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from the different angles and was converted to a physical surface value expressed in square mm by calibration. Experiments showed that the aboveground plant area measured this way correlates with the biomass of plant parts above ground.

In a first experiment, six lines in T1 generation were evaluated. There was an average increase of the harvest index and particularly two lines had a significant increase of 50% or more compared to the nullizygote lines (Table 2).

Table 2: Evaluation of the two best performing T1 events

harvestind x : 18.3					
Line	TR	null	dif	% dif	p-value
10	74.9	49.9	24.97	50	0.039
4	35	21.7	13.28	61	0.0656

Mean absolute values of the measurements of harvest index for the transgenic lines (TR) and control plants (null) in the T1 generation are given in columns 2 and 3, the absolute difference in column 4 and the difference in % in column 5, significance expressed as a p-value is given in column 6.

Example 5: Use of the invention in maize

The invention described herein can also be used in maize. To this aim, *GRUBX* is cloned under control of a seed-preferred promoter such as a prolamine promoter in a plant transformation vector suited for *Agrobacterium*-mediated corn transformation. Such vectors and methods for corn transformation have been described in literature (EP0604662, EP0672752, EP0971578, EP0955371, EP0558676, Ishida et al. 1996; Frame et al., 2002). Transgenic plants made by these methods are grown in the greenhouse for T1 seed production. Heritability is checked by progeny segregation analysis. Copy number of the transgene is checked by quantitative real-time PCR and/or Southern blot analysis. Expression levels of the transgene are determined by reverse PCR and/or Northern analysis. Transgenic lines with single copy insertions of the transgene and with varying levels of transgene expression are selected for T2 seed production through selfing or for crossing to different germplasm. Progeny seeds are germinated and grown in the field or in the greenhouse in conditions well adapted for maize (16:8 hr photoperiod, 26-28°C daytime and 22-24°C night time temperature) as well under water-deficient, nitrogen-deficient, and excess NaCl conditions. In the case of selfing, null segregants from the same parental line, as well as wild type plants of the same cultivar are used as controls. In the case of crossing, transgenics, null segregants and wild type plants of the same cultivar are crossed to a chosen parent and F1 plants from the transgenic cross are compared to F1 plants from the null segregant and the wild type crosses. The progeny plants resulting from the selfing or the crosses are evaluated on different biomass and growth parameters, including plant height, stem thickness, number of leaves, total above ground area, leaf greenness, time to maturity, flowering time, ear number, harvesting time. The seeds of these lines are also checked on various parameters, such as grain size, total grain yield per plant, and grain quality (starch content, protein content and oil content). Lines that are most significantly improved versus the controls for any of the above-

mentioned parameters are selected for further field testing and marker-assisted breeding, with the objective of transferring the field-validated transgenic traits into commercial germplasm. Methods for testing maize for growth and yield-related parameters in the field are well established in the art, as are techniques for introgressing specific loci (such as transgene containing loci) from one germplasm into another.

Claims

1. Method for modifying plant growth characteristics, comprising increasing expression in a plant of a nucleic acid sequence encoding a protein related to the human UBXD1 protein (GRUBX) and/or comprising increasing activity and/or increasing levels in a plant of a GRUBX protein.

2. Method of claim 1, wherein said GRUBX protein comprises at least a UBX and a PUG domain, and optionally also a Zinc finger domain.

3. Method according to claim 1or 2, wherein said increase is effected by recombinant means and/or chemical means.

4. Method according to claim 3, wherein said chemical means comprises the exogenous application of one or more compounds or elements capable of increasing expression in a plant of a GRUBX encoding nucleic acid and/or increasing activity and/or levels of a GRUBX protein in a plant.

5. Method according to any of claims 1 to 3, wherein said increasing expression comprises introduction into a plant of a nucleic acid capable of increasing activity and/or levels of a GRUBX protein and/or capable of increasing expression of a GRUBX gene, which nucleic acid is selected from:
   (i) Nucleic acids encoding transcription factors or activators of a GRUBX;
   (ii) GRUBX regulatory sequences; and
   (iii) Nucleic acids encoding ligands of a GRUBX protein.

6. Method according to any of claims 1 to 3, wherein said increasing expression comprises introducing a nucleic acid sequence encoding a GRUBX protein into a plant.

7. Method according to claim 6, wherein said nucleic acid is derived from a eukaryotic organism, preferably from a plant.

8. Method according to claim 7, wherein said nucleic acid is derived from a dicotyledonous plant, preferably from the family Solanaceae, more preferably from *Nicotiana tabacum*.

9. Method according to claim 7, wherein said nucleic acid is derived from a monocotyledonous plant, preferably from the family Poaceae, more preferably from *Oryza sativa*.

10. Method according to claim 8, wherein said nucleic acid is as represented by SEQ ID NO: 1 or is a portion thereof or is a sequence capable of hybridising therewith or encodes a GRUBX protein, wherein said GRUBX protein is represented by SEQ ID NO: 2 or a homologue, derivative or active fragment thereof.

11. Method according to any of claims 6 to 10, wherein said nucleic acid sequence and said proteins include variants chosen from:
    (i) an alternative splice variant of a nucleic acid sequence encoding a GRUBX protein or wherein said GRUBX protein is encoded by a splice variant;
    (ii) an allelic variant of a nucleic acid sequence encoding an GRUBX protein or wherein said GRUBX protein is encoded by an allelic variant;
    (iii) a nucleic acid sequence that is comprised on at least a part of an artificial chromosome, which artificial chromosome preferably also comprises one or more related gene family members;
    (iv) a functional portion of a GRUBX encoding nucleic acid;
    (v) sequence capable of hybridising to a GRUBX encoding nucleic acid;
    (vi) homologues, derivatives and active fragments of a GRUBX protein.

12. Method according to any of claims 1 to 11, wherein expression of said nucleic acid encoding a GRUBX protein is driven by a seed-preferred promoter, preferably the prolamine promoter.

13. Method according to any of claims 1 to 12, wherein said modified growth characteristic is selected from increased yield and/or modified plant architecture, each relative to corresponding wild type plants.

14. Method according to any of claims 1 to 12, wherein said modified yield and said modified plant architecture comprise at least increased harvest index relative to corresponding wild type plants.

15. Method for increasing the yield of a plant, which method comprises increasing expression in a plant of a GRUBX encoding nucleic acid and/or increasing activity and/or levels in a plant of a GRUBX protein.

16. Method for the production of a transgenic plant having modified growth characteristics, which method comprises:
    a. introducing into a plant or plant cell a nucleic acid sequence or a portion thereof encoding a GRUBX protein or a homologue, derivative or active fragment thereof;
    b. cultivating the plant cell under conditions promoting plant growth.

17. Method for the selection of plants having altered growth characteristics, which method is based on the selection of superior allelic variants of a GRUBX encoding sequence and which alleles give rise to altered growth characteristics in a plant.

18. Plants obtainable by a method according to any of claims 1 to 17.

19. An isolated nucleic acid sequence comprising:
    (i) a nucleic acid sequence represented by SEQ ID NO: 6, or the complement strand thereof;
    (ii) a nucleic acid sequence encoding an amino acid sequence represented by SEQ ID NO: 7, or homologues, derivatives or active fragments thereof;
    (iii) a nucleic acid sequence capable of hybridising (preferably under stringent conditions) with a nucleic acid sequence of (i) or (ii) above, which hybridising sequence preferably encodes a protein having GRUBX activity;
    (iv) a nucleic acid sequence according to (i) to (iii) above which is degenerate as a result of the genetic code;
    (v) a nucleic acid which is an allelic variant to the nucleic acid sequences according to (i) to (iv);
    (vi) a nucleic acid which is an alternative splice variant to the nucleic acid sequences according to (i) to (v);
    (vii) a nucleic acid sequence which has 75,00%, 80,00%, 85,00%, 90,00%, 95,00%, 96,00%, 97,00%, 98,00% or 99,00% sequence identity to any one or more of the sequence defined in (i) to (vi);
    (viii) a portion of a nucleic acid sequence according to any of (i) to (vii) above, which portion preferably encodes a protein having GRUBX activity,

20. An isolated protein comprising at least part of one of the polypeptides selected from the group consisting of:
    (i) a polypeptide as given in SEQ ID NO 4,

(ii) a polypeptide as given in SEQ ID NO 7

(iii) a polypeptide with an amino acid sequence which has at least 40.00% sequence identity, preferably 50.00%, 60.00%, 70.00% sequence identity, more preferable 80% or 90% sequence identity, most preferable 95.00%, 96.00%, 97.00%, 98.00% or 99.00% sequence identity to the amino acid sequence as given in SEQ ID NO 4 or SEQ ID NO 7,

(iv) a polypeptide comprising at least UBX domain and a PUG domain, and optionally a Zinc finger domain;

(v) a homologue, a derivative, an immunologically active and/or functional fragment of a protein as defined in any of (i) to (iv);

with the proviso that the protein sequence is not a sequence represented by SEQ ID NO 2, or database entries Q9ZU93, AAR01744, Q9D7L9, Q9BZV1, Q99PL6, ENSANGP00000020442, Q7SXA8, Q9V8K8, Q96IK9, ENSRNOP00000037228, or AAH07414..

21. Construct comprising:

(i) a nucleic acid sequence capable of increasing expression of a nucleic acid encoding a GRUBX protein and/or activity of a GRUBX protein;

(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

22. Construct according to claim 21, wherein said nucleic acid sequence capable of increasing expression of a nucleic acid encoding a GRUBX protein and/or activity of a GRUBX protein is a nucleic acid sequence encoding a protein according to any of (i) to (v) in claim 20.

23. Construct according to claim 22, wherein said nucleic acid sequence is represented by SEQ ID NO 1 or encodes a protein represented by SEQ ID NO 2.

24. Construct according to any of claims 21 to 23, wherein said control sequences comprise at least a seed-preferred promoter, preferably a prolamine promoter.

25. Construct according to any of claims 21 to 24, wherein said nucleic acid sequence capable of increasing expression of a nucleic acid encoding a GRUBX and/or activity of a GRUBX protein is oriented in sense direction relative to said control sequence.

26. Construct comprising an expression cassette essentially similar to SEQ ID NO 5.

27. Plant host cell having altered growth characteristics, characterized in that said host cell has increased expression of a nucleic acid sequence encoding a GRUBX protein and/or increased activity and/or levels of a GRUBX protein.

28. Transgenic plant having modified growth characteristics, characterized in that said plant has increased expression in a plant of a nucleic acid sequence encoding a GRUBX protein and/or increased activity in a plant of a GRUBX protein.

29. Transgenic plant according to claim 28, wherein said plant is a crop plant comprising soybean, sunflower, canola, alfalfa, rapeseed or cotton, preferably a monocotyledonous plant such as sugarcane, most preferably a cereal, such as rice, maize, wheat, millet, barley.

30. Plant cells, plant parts, including harvestable parts, propagules or progeny of a plant according to any of claims 18, 28 or 29.

31. Use of a nucleic acid sequence encoding a GRUBX protein, portions thereof or nucleic acids hybridising therewith, in modifying the growth characteristics of a plant.

32. Use of a GRUBX protein, homologues, derivatives and active fragments thereof, in modifying the growth characteristics of a plant.

33. Use of a GRUBX protein according to claim 20 in modifying the growth characteristics of a plant.

34. A composition comprising a GRUBX protein for use in modifying the growth characteristics of plants.

35. A composition comprising a nucleic acid encoding a GRUBX protein for use in modifying the growth characteristics of plants.

36. Use of a nucleic acid sequence in breeding programs, which nucleic acid sequence is capable of modulating expression of a nucleic acid encoding a GRUBX protein or capable of modulating the activity and/or levels of a GRUBX protein.

Abstract

Plants having modified growth characteristics and method for making the same

The present invention concerns a method for modifying the growth characteristics of plants by modulating expression in a plant of a nucleic acid sequence encoding a GRUBX protein and/or modulating activity and/or levels in a plant of a GRUBX protein. The invention also relates to transgenic plants having modified growth characteristics, which plants have modulated expression of a nucleic acid encoding a GRUBX protein.

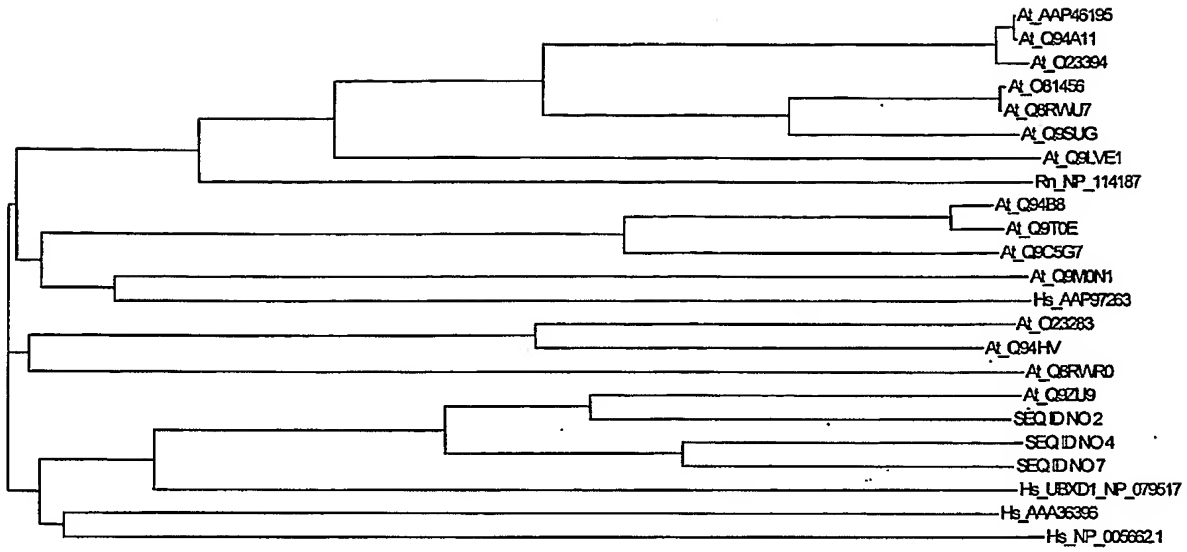

FIGURE 1 A

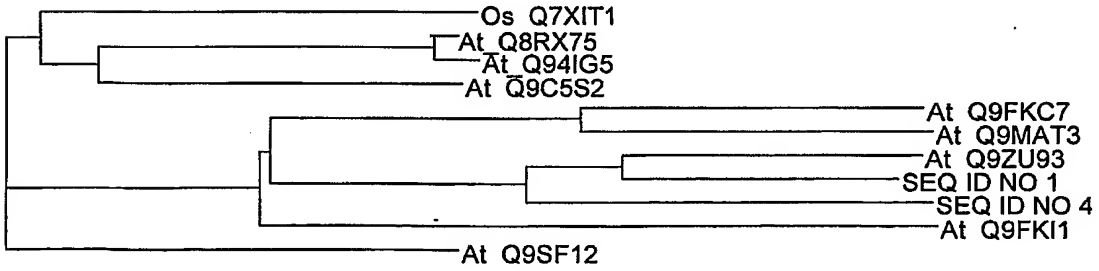

FIGURE 1B

```
Consensus terms for the UBX domain
YJE8_YEAST       KGFLAQ-NYCTLQLKLPNGYTISNTFP-PQTKLHKVRMWL----DYNCYDDGTP------
CONSENSUS/80%    s.t..t...splplRhssGpph.ppF..spp.lptlhthl........s.sss.......
CONSENSUS/65%    stpptt.stsplQlRhssGpphsppFs.uppslpplhphl....t...s.ssss......
CONSENSUS/50%    Ppppss.sps+LQlRhPDGpplsccFs.us-sLcplapaV....sptpssusss......

YJE8_YEAST        --------YLFHRNIPRVTLTRNDELKSLQEL-DLLPRSTLILEP
CONSENSUS/80%     ........apl.pshPch.hspt-..hph.p..thhspsslll..
CONSENSUS/65%     ........FpL.pshP+c.hsps-..ppl.ch.tlhPuslllt.
CONSENSUS/50%     ........FpLpssaPR+slocsD..KoLp-l.sLhPsuoLlLct


Consensus terms for the PUG domain
AAF74723         EASKLLLTYADNILRNPSDEKYRSIRIGNTAFSTRLLPVRGAVECLFEMGFEEG------
CONSENSUS/80%    tslt.LLphhcNhhpp.p-.........s..hpphlh.h.tsh..h....tF.ph......
CONSENSUS/65%    pslpsLL+hl+Nhhcc.pE.........stthpctltpl.sshphhhthtFscl.h....
CONSENSUS/50%    solccLLRhlRNlh+HhcEt.......tsppl+EplGslP-uhcpaFsssFPcLhh....

AAF74723          --ETHLIFPK
CONSENSUS/80%     ...phh.h.p
CONSENSUS/65%     ...phathhp
CONSENSUS/50%     ...csYplss
```

FIGURE 2 A

UBX domain present in SEQ ID NO 2

339 KAARKKYTKSIIRVQFPDGALLQGVFLPSEPTSALYEFVSAALKEPSLEFELLHPVLVKKRVIPHF
PAAGERAVTVEEEDLVPAALLKFK 428

PUG domain present in SEQ ID NO 2:

173 GSVEVVIKLLKNIVKEPENAKFRKIRMGNPKIKGAIGDVVGGVELLEFVGFELKEEGGEIWAVMDV
PS 240

UBX domain present in Q9ZU93

345 KAARKRYKRSMIRVQFPDGVVLQGVFAPWEPTFALYEFVSSALKEPSLQFELLDPVLVKRRVIPHT
PAPGQKPITLEDEELVPSALIRFRP 435

PUG domain present in Q9ZU93

181 SSIDVLLRLFKNIVKEPENAKFRKVRMSNAKIKEAIGDVAGGVELLELVGFELKEENDEIWAVMDV
PS 248

FIGURE 2 B

```
Q9ZU93          1 MDDVKDKLKGFMKKVNLSSSSGKFKGQGRVLGSSSSSSAPPVNPIQNRFN     50
                  |.|:|||:|||||||. ||||||||||||||||.||||.  |.|.: |.|:
SEQ_ID_NO_2     1 MGDMKDKVKGFMKKVT-SSSSGKFKGQGRVLGGSSSSG--PSNHV-NNFS     46

Q9ZU93         51 S-------SQAPNPTPRPKPNPNPLPEKPLSSSDQKISGSTR---NPDHD     90
                  |       .|.|:.|       ...|:|| |:|||:|.....   |....
SEQ_ID_NO_2    47 SHPLNTRQDQQPSYT-------KTSPQKP-SNSDQRIENICEIQFNKSES     88

Q9ZU93         91 PVRAPQDGFDPYGAFITSSNRSQNGYSL-SMFECPICKNPFTSEEEVSVH    139
                       :|||||:|..:||..|:..|||| ::||||:|.:.|.||||||.|
SEQ_ID_NO_2    89 -----KDGFDPFGELVTSGKRNPKGYSLTNVFECPVCGSGFVSEEEVSTH    133

Q9ZU93        140 VESCLG---DTN-GDESSFEKDNNDNDKSEMEKLVVVYLSGKPSESSIDV    185
                  ::|||.   .:| |.||..|.      |||:|..|..|:|||||||.|::|
SEQ_ID_NO_2    134 IDSCLSSEVSSNLGVESKVEV------KSELETCVSAYVSGKPSEGSVEV    177

Q9ZU93        186 LLRLFKNIVKEPENAKFRKVRMSNAKIKEAIGDVAGGVELLELVGFELKE    235
                  :::|.|||||||||||||:||.|.|||.|||||.|||||||.|||||||
SEQ_ID_NO_2    178 VIKLLKNIVKEPENAKFRKIRMGNPKIKGAIGDVVGGVELLEFVGFELKE    227

Q9ZU93        236 ENDEIWAVMDVPSEEQSILINKVVGYLEKRKTESSGSSAQV--MEPVAPK    283
                  |..|||||||||||||.:::..||..||.:|.|...|.:||  .|||.||
SEQ_ID_NO_2    228 EGGEIWAVMDVPSEEQLVMLKNVVSLLEPKKVEELASLSQVKASEPVEPK    277

Q9ZU93        284 KIDREIRVFFSVSENVASRIEVPDSFYSLSADEIKREADLRRKKIAESQL    333
                  ||||:|||||||.|:||:::||:||||::||.:|::|||::|:||:.:|:|
SEQ_ID_NO_2    278 KIDRQIRVFFSVPESVAAKIELPDSFFNLSREELRREAEMRKKKLEDSKL    327

Q9ZU93        334 LIPRSYKEKQAKAARKRYKRSMIRVQFPDGVVLQGVFAPWEPTFALYEFV    383
                  |||:||:|||||||||:|.:|:|||||||.:|||||.|.|||.||||||
SEQ_ID_NO_2    328 LIPKSYREKQAKAARKKYTKSIIRVQFPDGALLQGVFLPSEPTSALYEFV    377

Q9ZU93        384 SSALKEPSLQFELLDPVLVKRRVIPHTPAPGQKPITLEDEELVPSALIRF    433
                  |:|||||||:||||.|||||:|||||.||.|::.:|:|:|:|||:||:|
SEQ_ID_NO_2    378 SAALKEPSLEFELLHPVLVKKRVIPHFPAAGERAVTVEEEDLVPAALLKF    427

Q9ZU93        434 RPIETDSLVFTGLRNELLEISEPLS          458
                  :||||||:|||||.||||||||||.
SEQ_ID_NO_2    428 KPIETDSVVFTGLCNELLEISEPLETGSVASS    459
```

FIGURE 2 C

```
SEQ_ID_NO2       1 MGDMKDKVKGFMKKVTSSSSG---KFKGQGRVLGGSSSSGPSNHVNNFSS     47
                   .||||:|.|||||||||.||   .|||...|||    ||||..    ||
SEQ_ID_NO4_C     1   mmkdkmkefmkkvtssgsgtpssfkgtshvlg----sgpsps----ss     40

SEQ_ID_NO2      48 HPL----NTRQDQQPSYTKTSPQKPSNSDQRIENICEIQFNKSESKDGFD     93
                   ||.    |...:.:|:..:|||..|......:.:
SEQ_ID_NO4_C    41 hpaarssnpspnlrpapkrtspptpptlttdlts----------------     74

SEQ_ID_NO2      94 PFGELVTSGKRNPKGYS---LTNVFECPVCGSGFVSEEEVSTHIDSCLSS    140
                    |..||....|.|....   ......||.||..|.||..||.|:|.||:|
SEQ_ID_NO4_C    75 -ftplvcyssrrpdangtaCavatvacpscgdafpselavsehldgclas    123

SEQ_ID_NO2     141 EVSSNLGVESKVEVKSELETCVSAYVSG--KPSEGSVEVVIKLLKNIVKE    188
                   ..    |..::          .:||::.  .|...|||||.:||.|:::|
SEQ_ID_NO4_C   124 ag----garar----------aaaylaadppppaasvevvkrllgnllre    159

SEQ_ID_NO2     189 PENAKFRKIRMGNPKIKGAIGDVVGGVELLEFVGFELKEEGGEIWAVMD-    237
                   |.|.|||::|:|||:||.|:.|..|||||||.|||.:.:|||||.:||||
SEQ_ID_NO4_C   160 pgndkfrrvrlgnprikealadrdggvelleavgftvgdeggepfavmde    209

SEQ_ID_NO2     238 VPSEEQLVMLKNVVSLLE-------PKKVEELASLSQVKASEPVE-PKKI    279
                   |||:.:|..:.:..|.|||       |.|.|..|..|....|:..| .|.|
SEQ_ID_NO4_C   210 vpsdprlngirravlllegahpsappvkaeaeakescsnvsdvqegakti    259

SEQ_ID_NO2     280 DRQIRVFFSVPESVAAKIELPDSFFNLSREELRREAEMRKKKLEDSKLLI    329
                   |||||||.|||.|..|:.::||||:.||.||:|.||:||:::||.|:|||
SEQ_ID_NO4_C   260 drqirvfvsvpgssmaqndvpdsfyklsgeeirneakmrrerleqsrlli    309

SEQ_ID_NO2     330 PKSYREKQAKAARKKYTKSIIRVQFPDGALLQGVFLPSEPTSALYEFVSA    379
                   ||||:||||.|||:||.:::|||||||..:|||:|||.|.||:||||||::
SEQ_ID_NO4_C   310 pksykekqalaarqkykqavirvqfpdrmilqgiflpgeatsslyefvts    359

SEQ_ID_NO2     380 ALKEPSLEFELLHPVLVKKRVIPHFPAAGERAVTVEEEDLVPAALLKFKP    429
                   |||:..|||||:.|.:.|.||:||||..||||.|::||:|||:|||||.|
SEQ_ID_NO4_C   360 alkqsglefelispaipkprvvphfpnpgerartlqeeelvpsallkfip    409

SEQ_ID_NO2     430 IETDSVVFTGLCNELLEISEPLETGSVASS    459
                   .||||:|||||.:|||..||||...|.
SEQ_ID_NO4_C   410 ketdsmvftglldellmaseplpaasq      436
```

FIGURE 2 D

```
SEQ_ID_NO_4       1 mmkdkmkefmkkvt--ssgsgtpssfkgtshvlgsgpspssshpaarssn     48
                    |||:|||:.|:||| ||.|.:.|||||:|||||||.|||     |.||
SEQ_ID_NO_7       1 MMKEKMKDLMRKVTSSSSSSSSSSSFKGTAHVLGSGPDPSS-----RPSN     45

SEQ_ID_NO_4      49 pspnlrpapkrtspptpptlttdltsftplvcyssrr--pdangtaCava     96
                    |:|:...||:|.:..:....::....::||:..||||  |.|...|....
SEQ_ID_NO_7      46 PTPSRPAAPRREAAASARPPSSGFAPYSPLISTSSRRTDPPAGAGAGEDD     95

SEQ_ID_NO_4      97 tvacpscgdafpselavsehldgclasaggararaaaylaadppppaasv    146
                    .||||||.:.|||||||:|||||||:|||||.|||||||.|  |||::|
SEQ_ID_NO_7      96 AVACPSCAEPFPSELAVSDHLDGCLAAAGGARPRAAAYLAGD--PPASAV    143

SEQ_ID_NO_4     147 evvkrllgnllrepgndkfrrvrlgnprikealadrdggvelleavgftv    196
                    ||||||||||||.:|.|||:|:|||||||||||||||:|||:|||||||.|
SEQ_ID_NO_7     144 EVVKRLLGNLLSDPRNDKYRKVRLGNPRIKEALADREGGVDLLEAVGFRV    193

SEQ_ID_NO_4     197 gdeggepfavmdevpsdprlngirravlllegahpsappvkaeaeakesc    246
                    .|||||.||:|||||.|.||.|||:||||||.|.||.|| :.:|:|||:|
SEQ_ID_NO_7     194 ADEGGELFALMDEVPGDARLGGIRQAVLLLERARPSTPP-QTQADAKETC    242

SEQ_ID_NO_4     247 sn-vsdvqegaktidrqirvfvsvpgssmaqndvpdsfyklsgeeirnea    295
                    .| ||:.|...|.:|||||||.||..||:|:||:|||||.||.|||||||
SEQ_ID_NO_7     243 PNGVSEEQGIKKPVDRQIRVFFSVAASSVAENDLPDSFYSLSNEEIRNEA    292

SEQ_ID_NO_4     296 kmrrerleqsrllipksykekqalaarqkykqavirvqfpdrmilqgifl    345
                    |||||||||||||||||||||||||||||||||||:||:||||.:||||:||
SEQ_ID_NO_7     293 KMRRERLEQSRLLIPKSYKEKQALAARQKYKQALIRIQFPDGVILQGVFL    342

SEQ_ID_NO_4     346 pgeatsslyefvtsalkqsglefelispaipkprvvphfpnpgerartlq    395
                    |.|..|||||||.|:|||..|||:||.||.|:.||:|.||.|||:|||||:
SEQ_ID_NO_7     343 PAEPISSLYEFVASSLKQPSLEFDLICPAGPRTRVIPPFPKPGEQARTLR    392

SEQ_ID_NO_4     396 eeelvpsallkfipketdsmvftglldellmaseplpaasq    436
                    :|:|||||.|.|.||||||:||||||||||..|||..:||.
SEQ_ID_NO_7     393 DEDLVPSARLTFKPKETDSVVFTGLLDELLETSEPFTSASS    433
```

FIGURE 2 E

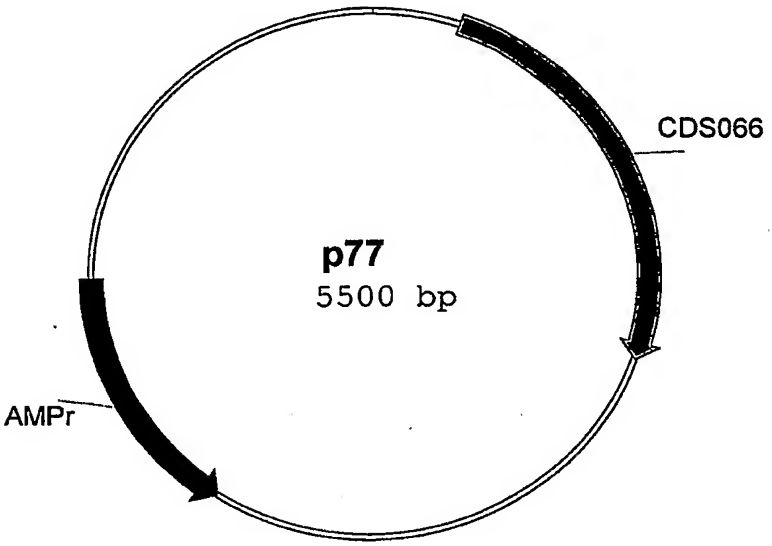

FIGURE 3

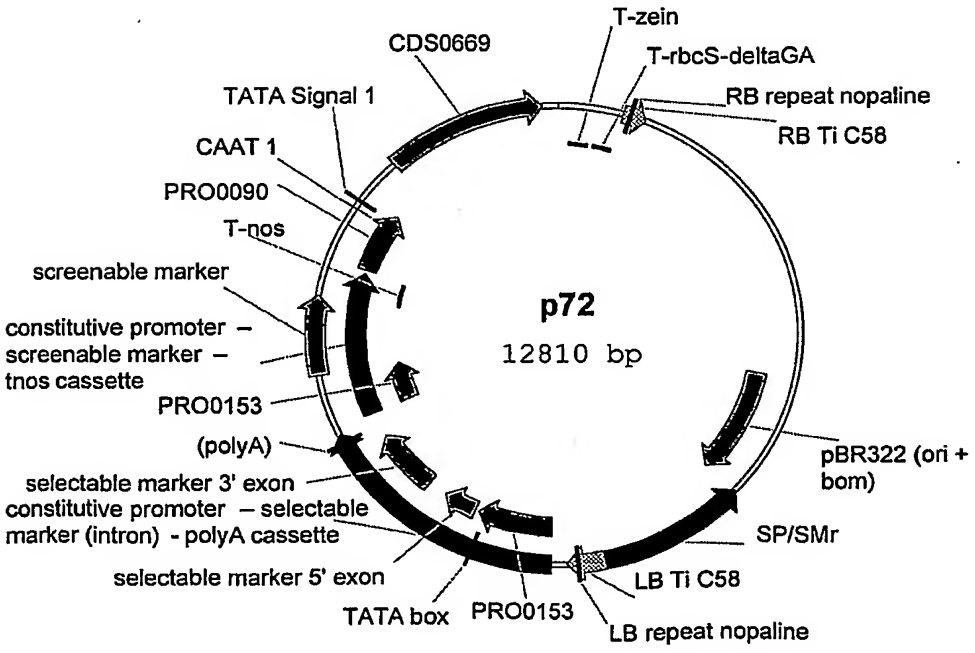

FIGURE 4

SEQ ID NO 1: GRUBX coding sequence

atgggtgacatgaaggataaagtcaaagggttcatgaaaaaagtcacatcttcttcttcaggtaagttta
aaggccaaggtagggttttgggtggttcatcttcttcaggaccctcaaatcatgtcaataatttttcatc
acatcccctaaatacaaggcaagatcaacaaccttcatatacaaaaacttcgcctcaaaaaccaagtaat
tctgatcaaagaattgagaatatatgtgaaattcagttcaacaaaagtgaatcaaaggatggttttgatc
catttggtgaattagtcacttctgggaagagaaacccaaaagggtattcacttactaatgtgtttgaatg
ccctgtctgtggtagtggttttgtttctgaagaagaggtgtcaactcatattgatagctgtttaagttct
gaagtgtcttctaatttgggagttgaaagtaaagttgaagttaaaagtgaattggaaacatgtgttagtg
catatgtttcagggaagccctcagaagggtcagttgaagtggtcattaagttgttaaagaatattgtgaa
ggaaccagagaatgccaagtttaggaaaataaggatggggaatccaaaaataaaaggtgctataggtgat
gttgtaggaggagtggagctattggaatttgttggatttgagttgaaagaagaaggtggggaaatttggg
ctgtgatggatgttccttctgaagaacaacttgttatgcttaagaatgtagtttcactcttggaaccgaa
gaaggttgaagagttggcgtccttatcccaagttaaggcgagtgaaccagttgagccgaagaagattgat
agacagattcgagtgttctttttctgttcccgagagcgtagcagcaaaaattgagctacctgattccttct
ttaacctctcacgtgaggaattgagaagagaagcagagatgaggaagaagaaattagaagattccaaatt
attgattcctaaatcttatcgggaaaagcaggcaaaagctgcaagaaagaagtacacaaaatccattatc
cgtgtacagtttccagatggagcattgcttcaaggtgtctttctaccttcggagccaactagtgctcttt
atgagtttgtgagcgcagcgttaaaggaaccaagcttagagttcgaattgttacatccggtgcttgttaa
aaagcgggtgattccccatttccagctgctggggagagggctgtaacagttgaagaggaggatttggtt
cctgcagctctactcaaatttaaacctatcgaaacagattctgttgtttttactggtctttgtaatgagc
ttcttgaaattagcgagcccctcgagaccggatcagttgcttcctcgtaa

SEQ ID NO 2: GRUBX, deduced protein sequence

MGDMKDKVKGFMKKVTSSSSGKFKGQGRVLGGSSSSGPSNHVNNFSSHPLNTRQDQQPSYTKTSPQKPSN
SDQRIENICEIQFNKSESKDGFDPFGELVTSGKRNPKGYSLTNVFECPVCGSGFVSEEEVSTHIDSCLSS
EVSSNLGVESKVEVKSELETCVSAYVSGKPSEGSVEVVIKLLKNIVKEPENAKFRKIRMGNPKIKGAIGD
VVGGVELLEFVGFELKEEGGEIWAVMDVPSEEQLVMLKNVVSLLEPKKVEELASLSQVKASEPVEPKKID
RQIRVFFSVPESVAAKIELPDSFFNLSREELRREAEMRKKKLEDSKLLIPKSYREKQAKAARKKYTKSII
RVQFPDGALLQGVFLPSEPTSALYEFVSAALKEPSLEFELLHPVLVKKRVIPHFPAAGERAVTVEEEDLV
PAALLKFKPIETDSVVFTGLCNELLEISEPLETGSVASS

SEQ ID NO 3: *GRUBX* orthologue from sugarcane

atgatgaaggacaagatgaaggagttcatgaagaaggtcacctcctccgggtccgggacccccctcctcct
tcaagggcacctcccacgtcctcggctccggcccctcccctcctcctcccaccccgctgcccgctcctc
aaaccctagcccaaacctcaggcccgctcctaagcggacctcgccacctaccccgcccactttaaccacc
gatttgacctccttcacgcccctcgtctgctactcctcccgccgccccgacgcgaacggcaccgcgnnng
ccgtcgccaccgtcgcgtgccccagctgcggagacgcgtttccgtccgagctcgccgtctccgagcatct
cgacggctgcctcgcgtcggcgggggcgcccgcgcgcgccgccgcgtacctcgccgccgacccgcct
ccgcccgcggcctccgtagaggtggtcaaacgcctgctgggcaacctgctccgggagcccggcaacgata
agttcaggcgggtgagattgggtaacccgcggatcaaggaggccctggcagacagggatggcggggtgga
gctcctggaggccgtcggcttcacagttggggatgagggcggggagcccttcgccgtgatggacgaagtg
cctagcgaccctaggctcaacgggatcaggagggccgtcctcctgctcgaggggcacacccctctgcgc
ctccagtgaaggcggaggctgaggccaaggagagctgcagcaatgtgtctgacgtgcaggagggtgctaa
gactattgatcggcagattcgggtatttgtctctgttcctgggagttctatggcacaaaatgatgtacca
gattctttttacaagcttagtggtgaggagataaggaatgaagcaaagatgaggagggaaaggttagaac
aatctcgattgctgataccaaagtcttacaaggagaaacaggcattggctgctcgacagaagtataaaca
agcagtcattcgagttcagtttccagatagaatgattcttcagggcatattcctaccaggagaggccact
agttcactgtatgagttcgtcacatctgctctgaagcaatcaggtttggaattcgaacttatctctccag
ccatacctaagccacgtgtggtgccccatttccaaacccgggagagcgggcacgcaccttgcaagagga
ggagctggtcccatctgcgctcctcaagttcattcccaaggagactgattccatggttttcaccggtttg
cttgatgagcttctcatggccagtgagccgcttcctgctgcatcacaatga

FIGURE 5

SEQ ID NO 4: sugarcane GRUBX orthologue, deduced protein sequence

MMKDKMKEFMKKVTSSGSGTPSSFKGTSHVLGSGPSPSSSHPAARSSNPSPNLRPAPKRTSPPTPPTLTT
DLTSFTPLVCYSSRRPDANGTAXAVATVACPSCGDAFPSELAVSEHLDGCLASAGGARARAAAYLAADPP
PPAASVEVVKRLLGNLLREPGNDKFRRVRLGNPRIKEALADRDGGVELLEAVGFTVGDEGGEPFAVMDEV
PSDPRLNGIRRAVLLLEGAHPSAPPVKAEAEAKESCSNVSDVQEGAKTIDRQIRVFVSVPGSSMAQNDVP
DSFYKLSGEEIRNEAKMRRERLEQSRLLIPKSYKEKQALAARQKYKQAVIRVQFPDRMILQGIFLPGEAT
SSLYEFVTSALKQSGLEFELISPAIPKPRVVPHFPNPGERARTLQEEELVPSALLKFIPKETDSMVFTGL
LDELLMASEPLPAASQ

SEQ ID NO 5: expression cassette comprising *GRUBX* (1011-2390) operably linked to the prolamine promoter (1-654) and the T-Zein + T-Rubisco deltaG terminator (2615-2808 and 2852-3048):

cttctacatcggcttaggtgtagcaacacgactttattattattattattattattattattttaca
aaaatataaaatagatcagtccctcaccacaagtagagcaagttggtgagttattgtaaagttctacaaa
gctaatttaaaagttattgcattaacttatttcatattacaaacaagagtgtcaatggaacaatgaaaac
catatgacatactataattttgtttttattattgaaattatataattcaaagagaataaatccacatagc
cgtaaagttctacatgtggtgcattaccaaaatatatatagcttacaaaacatgacaagcttagtttgaa
aaattgcaatccttatcacattgacacataaagtgagtgatgagtcataatattattttctttgctaccc
atcatgtatatatgatagccacaaagttactttgatgatgatatcaaagaacatttttaggtgcacctaa
cagaatatccaaataatatgactcacttagatcataatagagcatcaagtaaaactaacactctaaagca
accgatgggaaagcatctataaatagacaagcacaatgaaaatcctcatcatccttcaccacaattcaaa
tattatagtttgaagcatagtagtaatttaaatcaactagggatatcacaagtttgtacaaaaaagcaggc
tggtaccggtccggaattcccgggatatcgtcgacccacgcgtccgccaatatcagatttctttcatgaa
ctccacttccaatttctcattgcttcttcttcccatttccacctccaaagccatccttccagaaaacctt
gttccttacatttcttagccccaaaaaagattcccatctcaattccacaaaaaaacacaaggagatctaa
ggaaattccccgcctctatatatagaggtggaattgttcctgaatttggtttgaattgattgattgac
agattttggtgagagggtgttattgaaaaaatgggtgacatgaaggataaagtcaaagggttcatgaaaa
aagtcacatcttcttcttcaggtaagtttaaaggccaaggtagggttttgggtggttcatcttcttcagg
accctcaaatcatgtcaataatttttcatcacatcccctaaatacaaggcaagatcaacaaccttcatat
acaaaaacttcgcctcaaaaaccaagtaattctgatcaaagaattgagaatatatgtgaaattcagttca
acaaaagtgaatcaaaggatggttttgatccatttggtgaattagtcacttctgggaagagaaacccaaa
agggtattcacttactaatgtgtttgaatgccctgtctgtggtagtggttttgtttctgaagaagaggtg
tcaactcatattgatagctgtttaagttctgaagtgtcttctaatttgggagttgaaagtaaagttgaag
ttaaaagtgaattggaaacatgtgttagtgcatatgtttcagggaagccctcagaagggtcagttgaagt
ggtcattaagttgttaaagaatattgtgaaggaaccagagaatgccaagtttaggaaaataaggatgggg
aatccaaaaataaaaggtgctataggtgatgttgtaggaggagtggagctattggaatttgttggatttg
agttgaaagaagaaggtggggaaatttgggctgtgatggatgttccttctgaagaacaacttgttatgct
taagaatgtagtttcactcttggaaccgaagaaggttgaagagttggcgtccttatcccaagttaaggcg
agtgaaccagttgagccgaagaagattgatagacagattcgagtgttcttttctgttcccgagagcgtag
cagcaaaaattgagctacctgattccttctttaacctctcacgtgaggaattgagaagagaagcagagat
gaggaagaagaaattagaagattccaaattattgattcctaaatcttatcgggaaaagcaggcaaaagct
gcaagaaagaagtacacaaaatccattatccgtgtacagtttccagatggagcattgcttcaaggtgtct
ttctaccttcggagccaactagtgctctttatgagtttgtgagcgcagcgttaaaggaaccaagcttaga
gttcgaattgttacatccggtgcttgttaaaaagcgggtgattccccatttccagctgctggggagagg
gctgtaacagttgaagaggaggatttggttcctgcagctctactcaaatttaaacctatcgaaacagatt
ctgttgtttttactggtctttgtaatgagcttcttgaaattagcgagcccctcgagaccggatcagttgc
ttcctcgtaagctctaaattacatcagactttgaattcttctgagtgttggaaaccttataaaactctct
gcgccgggaatgctgcggccgctctagagtatccctcgaggggcccaagcttacgcgtacccagctttct
tgtacaaagtggtgatatcacaagcccgggcggtcttctagggataacagggtaattatatccctctaga
tcacaagcccgggcggtcttctacgatgattgagtaataatgtgtcacgcatcaccatgggtggcagtgt

FIGURE 5 (continued)

cagtgtgagcaatgacctgaatgaacaattgaaatgaaaagaaaaaaagtactccatctgttccaaatta
aaattcattttaacctttaataggtttatacaataattgatatatgttttctgtatatgtctaatttgt
tatcatccgggcggtcttctagggataacagggtaattatatccctctagacaacacacaacaaataaga
gaaaaaacaaataatattaatttgagaatgaacaaaaggaccatatcattcattaactcttctccatcca
tttccatttcacagttcgatagcgaaaaccgaataaaaaacacagtaaattacaagcacaacaaatggta
caagaaaaacagttttcccaatgccataatactcgaac

SEQ ID NO 6: *Oryza sativa GRUBX* orthologue, DNA sequence
atgatgaaggaaaagatgaaggatctcatgaggaaggtcacctcctcctcctcctcctcgtcgtcct
cctccttcaagggcaccgcccacgtcctcggctccggccccgacccctcctcccgcccctccaaccctac
ccctagtcgccccgctgccccccggcgagaggccgccgcctccgcgaggccgccctcctccggcttcgcc
ccctactccccgctcatctccacctcctcccgccgcaccgacccacccgcgggggcgggggcgggggagg
acgacgccgtcgcgtgccccagctgcgccgagccgttcccctccgagctggcggtgtcggaccacctcga
cggctgcctcgcggcggcgggggggagcccgcccccgcgcggccgcctacctggccggcgacccccccgcg
tccgccgtggaggtggtgaagaggctgctcgggaacctgctctccgacccccggaacgacaagtacagga
aggtcaggctcgggaacccgaggatcaaggaggccctggcggacagggagggcggggtggatctcctcga
ggccgtggggttcagggtcgccgacgagggcggggagctcttcgccctcatggacgaggtgcccggggac
gcgaggctcggcggcatcaggcaggccgtgctcctgctcgagagggcccggccatcgacgccgccgcaga
cacaggcagatgccaaagagacttgcccgaatggagttagcgaagagcaggggattaagaagccggttga
tcgtcagattcgggtgttcttctctgttgctgcaagttctgttgcagaaaatgatctaccagattctttc
tatagccttagtaatgaggagatcaggaatgaggcaaagatgaggagggagaggctagaacaatctcggt
tgttgattccaaagtcatataaggagaagcaggcactggctgcccgacagaagtataaacaagctctgat
tcgaattcagtttccggatggagtaattctgcagggtgtgttccttcccgcggagcccattagttcacta
tatgagtttgtcgcatcttctttgaagcagccaagtttggaattcgatcttatctgtccagcgggtccta
ggacacgcgtcatacctccttttccaaaaccaggggaacaggcacgcacactgcgggatgaagacctagt
cccttctgctcggctcaccttcaagcccaaagagactgattcggttgtgttcactggtctgttggacgag
cttcttgagaccagtgagccgttcacatctgcatcctcatga

SEQ ID NO 7: *Oryza sativa GRUBX* orthologue, deduced protein sequence
MMKEKMKDLMRKVTSSSSSSSSSSFKGTAHVLGSGPDPSSRPSNPTPSRPAAPRREAAASARPPSSGFA
PYSPLISTSSRRTDPPAGAGAGEDDAVACPSCAEPFPSELAVSDHLDGCLAAAGGARPRAAAYLAGDPPA
SAVEVVKRLLGNLLSDPRNDKYRKVRLGNPRIKEALADREGGVDLLEAVGFRVADEGGELFALMDEVPGD
ARLGGIRQAVLLLERARPSTPPQTQADAKETCPNGVSEEQGIKKPVDRQIRVFFSVAASSVAENDLPDSF
YSLSNEEIRNEAKMRRERLEQSRLLIPKSYKEKQALAARQKYKQALIRIQFPDGVILQGVFLPAEPISSL
YEFVASSLKQPSLEFDLICPAGPRTRVIPPFPKPGEQARTLRDEDLVPSARLTFKPKETDSVVFTGLLDE
LLETSEPFTSASS

FIGURE 5 (continued)

SEQUENCE LISTING

<110> CropDesign N.V.

<120> Plants having modified growth characteristics and method for making the same

<130> CD-107-Prio

<160> 7

<170> PatentIn version 3.1

<210> 1
<211> 1380
<212> DNA
<213> Nicotiana tabacum

<400> 1
atgggtgaca tgaaggataa agtcaaaggg ttcatgaaaa aagtcacatc ttcttcttca 60
ggtaagttta aaggccaagg tagggttttg ggtggttcat cttcttcagg accctcaaat 120
catgtcaata atttttcatc acatccccta aatacaaggc aagatcaaca accttcatat 180
acaaaaactt cgcctcaaaa accaagtaat tctgatcaaa gaattgagaa tatatgtgaa 240
attcagttca acaaaagtga atcaaaggat ggttttgatc catttggtga attagtcact 300
tctgggaaga gaaacccaaa agggtattca cttactaatg tgtttgaatg ccctgtctgt 360
ggtagtggtt ttgtttctga agaagaggtg tcaactcata ttgatagctg tttaagttct 420
gaagtgtctt ctaatttggg agttgaaagt aaagttgaag ttaaaagtga attggaaaca 480
tgtgttagtg catatgtttc agggaagccc tcagaagggt cagttgaagt ggtcattaag 540
ttgttaaaga atattgtgaa ggaaccagag aatgccaagt ttaggaaaat aaggatgggg 600
aatccaaaaa taaaaggtgc tataggtgat gttgtaggag gagtggagct attggaattt 660
gttggatttg agttgaaaga agaaggtggg gaaatttggg ctgtgatgga tgttccttct 720
gaagaacaac ttgttatgct taagaatgta gtttcactct tggaaccgaa gaaggttgaa 780
gagttggcgt ccttatccca agttaaggcg agtgaaccag ttgagccgaa gaagattgat 840
agacagattc gagtgttctt ttctgttccc gagagcgtag cagcaaaaat tgagctacct 900
gattccttct ttaacctctc acgtgaggaa ttgagaagag aagcagagat gaggaagaag 960
aaattagaag attccaaatt attgattcct aaatcttatc gggaaaagca ggcaaaagct 1020
gcaagaaaga agtacacaaa atccattatc cgtgtacagt ttccagatgg agcattgctt 1080
caaggtgtct ttctaccttc ggagccaact agtgctcttt atgagtttgt gagcgcagcg 1140
ttaaaggaac caagcttaga gttcgaattg ttacatccgg tgcttgttaa aaagcgggtg 1200
attccccatt ttccagctgc tggggagagg gctgtaacag ttgaagagga ggatttggtt 1260
cctgcagctc tactcaaatt taaacctatc gaaacagatt ctgttgtttt tactggtctt 1320
tgtaatgagc ttcttgaaat tagcgagccc ctcgagaccg gatcagttgc ttcctcgtaa 1380

<210> 2
<211> 459
<212> PRT
<213> Nicotiana tabacum

<400> 2
Met Gly Asp Met Lys Asp Lys Val Lys Gly Phe Met Lys Lys Val Thr
1 5 10 15

Ser Ser Ser Ser Gly Lys Phe Lys Gly Gln Gly Arg Val Leu Gly Gly
20 25 30

Ser Ser Ser Ser Gly Pro Ser Asn His Val Asn Asn Phe Ser Ser His
35 40 45

Pro Leu Asn Thr Arg Gln Asp Gln Gln Pro Ser Tyr Thr Lys Thr Ser

```
    50                  55                  60

Pro Gln Lys Pro Ser Asn Ser Asp Gln Arg Ile Glu Asn Ile Cys Glu
65                  70                  75                  80

Ile Gln Phe Asn Lys Ser Glu Ser Lys Asp Gly Phe Asp Pro Phe Gly
                85                  90                  95

Glu Leu Val Thr Ser Gly Lys Arg Asn Pro Lys Gly Tyr Ser Leu Thr
            100                 105                 110

Asn Val Phe Glu Cys Pro Val Cys Gly Ser Gly Phe Val Ser Glu Glu
        115                 120                 125

Glu Val Ser Thr His Ile Asp Ser Cys Leu Ser Ser Glu Val Ser Ser
    130                 135                 140

Asn Leu Gly Val Glu Ser Lys Val Glu Val Lys Ser Glu Leu Glu Thr
145                 150                 155                 160

Cys Val Ser Ala Tyr Val Ser Gly Lys Pro Ser Glu Gly Ser Val Glu
                165                 170                 175

Val Val Ile Lys Leu Leu Lys Asn Ile Val Lys Glu Pro Glu Asn Ala
            180                 185                 190

Lys Phe Arg Lys Ile Arg Met Gly Asn Pro Lys Ile Lys Gly Ala Ile
        195                 200                 205

Gly Asp Val Val Gly Gly Val Glu Leu Leu Glu Phe Val Gly Phe Glu
    210                 215                 220

Leu Lys Glu Glu Gly Gly Glu Ile Trp Ala Val Met Asp Val Pro Ser
225                 230                 235                 240

Glu Glu Gln Leu Val Met Leu Lys Asn Val Val Ser Leu Leu Glu Pro
                245                 250                 255

Lys Lys Val Glu Glu Leu Ala Ser Leu Ser Gln Val Lys Ala Ser Glu
            260                 265                 270

Pro Val Glu Pro Lys Lys Ile Asp Arg Gln Ile Arg Val Phe Phe Ser
        275                 280                 285

Val Pro Glu Ser Val Ala Ala Lys Ile Glu Leu Pro Asp Ser Phe Phe
    290                 295                 300

Asn Leu Ser Arg Glu Glu Leu Arg Arg Glu Ala Glu Met Arg Lys Lys
305                 310                 315                 320

Lys Leu Glu Asp Ser Lys Leu Leu Ile Pro Lys Ser Tyr Arg Glu Lys
                325                 330                 335

Gln Ala Lys Ala Ala Arg Lys Lys Tyr Thr Lys Ser Ile Ile Arg Val
            340                 345                 350

Gln Phe Pro Asp Gly Ala Leu Leu Gln Gly Val Phe Leu Pro Ser Glu
        355                 360                 365
```

```
Pro Thr Ser Ala Leu Tyr Glu Phe Val Ser Ala Ala Leu Lys Glu Pro
    370                 375                 380

Ser Leu Glu Phe Glu Leu Leu His Pro Val Leu Val Lys Lys Arg Val
385                 390                 395                 400

Ile Pro His Phe Pro Ala Ala Gly Glu Arg Ala Val Thr Val Glu Glu
                405                 410                 415

Glu Asp Leu Val Pro Ala Ala Leu Leu Lys Phe Lys Pro Ile Glu Thr
            420                 425                 430

Asp Ser Val Val Phe Thr Gly Leu Cys Asn Glu Leu Leu Glu Ile Ser
        435                 440                 445

Glu Pro Leu Glu Thr Gly Ser Val Ala Ser Ser
    450                 455
```

<210> 3
<211> 1311
<212> DNA
<213> Saccharum officinarum

<220>
<221> misc_feature
<222> (277)..(279)
<223> n can be any nucleotide

<400> 3

```
atgatgaagg acaagatgaa ggagttcatg aagaaggtca cctcctccgg gtccgggacc     60
ccctcctcct tcaagggcac ctcccacgtc ctcggctccg gcccctcccc ctcctcctcc    120
caccccgctg cccgctcctc aaaccctagc ccaaacctca ggcccgctcc taagcggacc    180
tcgccaccta ccccgcccac tttaaccacc gatttgacct ccttcacgcc cctcgtctgc    240
tactcctccc gccgccccga cgcgaacggc accgcgnnng ccgtcgccac cgtcgcgtgc    300
cccagctgcg gagacgcgtt tccgtccgag ctcgccgtct ccgagcatct cgacggctgc    360
ctcgcgtcgg cgggggggcgc ccgcgcgcgc gccgccgcgt acctcgccgc cgacccgcct    420
ccgccccgcgg cctccgtaga ggtggtcaaa cgcctgctgg gcaacctgct ccgggagccc    480
ggcaacgata agttcaggcg ggtgagattg ggtaacccgc ggatcaagga ggccctggca    540
gacagggatg gcggggtgga gctcctggag gccgtcggct tcacagttgg ggatgagggc    600
ggggagccct tcgccgtgat ggacgaagtg cctagcgacc ctaggctcaa cgggatcagg    660
agggccgtcc tcctgctcga gggggcacac ccctctgcgc ctccagtgaa ggcggaggct    720
gaggccaagg agagctgcag caatgtgtct gacgtgcagg agggtgctaa gactattgat    780
cggcagattc gggtatttgt ctctgttcct gggagttcta tggcacaaaa tgatgtacca    840
gattcttttt acaagcttag tggtgaggag ataaggaatg aagcaaagat gaggagggaa    900
aggttagaac aatctcgatt gctgatacca aagtcttaca aggagaaaca ggcattggct    960
gctcgacaga agtataaaca agcagtcatt cgagttcagt ttccagatag aatgattctt   1020
cagggcatat tcctaccagg agaggccact agttcactgt atgagttcgt cacatctgct   1080
ctgaagcaat caggtttgga attcgaactt atctctccag ccatacctaa gccacgtgtg   1140
gtgccccatt ttccaaaccc gggagagcgg gcacgcacct tgcaagagga ggagctggtc   1200
ccatctgcgc tcctcaagtt cattcccaag gagactgatt ccatggtttt caccggtttg   1260
cttgatgagc ttctcatggc cagtgagccg cttcctgctg catcacaatg a             1311
```

<210> 4
<211> 436
<212> PRT
<213> Saccharum officinarum

<220>
<221> MISC_FEATURE

<222> (93)..(93)
<223> Xaa can be any amino acid

<400> 4

```
Met Met Lys Asp Lys Met Lys Glu Phe Met Lys Lys Val Thr Ser Ser
1               5                  10                  15

Gly Ser Gly Thr Pro Ser Ser Phe Lys Gly Thr Ser His Val Leu Gly
            20                  25                  30

Ser Gly Pro Ser Pro Ser Ser Ser His Pro Ala Ala Arg Ser Ser Asn
        35                  40                  45

Pro Ser Pro Asn Leu Arg Pro Ala Pro Lys Arg Thr Ser Pro Pro Thr
    50                  55                  60

Pro Pro Thr Leu Thr Thr Asp Leu Thr Ser Phe Thr Pro Leu Val Cys
65                  70                  75                  80

Tyr Ser Ser Arg Arg Pro Asp Ala Asn Gly Thr Ala Xaa Ala Val Ala
                85                  90                  95

Thr Val Ala Cys Pro Ser Cys Gly Asp Ala Phe Pro Ser Glu Leu Ala
            100                 105                 110

Val Ser Glu His Leu Asp Gly Cys Leu Ala Ser Ala Gly Gly Ala Arg
        115                 120                 125

Ala Arg Ala Ala Ala Tyr Leu Ala Ala Asp Pro Pro Pro Pro Ala Ala
    130                 135                 140

Ser Val Glu Val Val Lys Arg Leu Leu Gly Asn Leu Leu Arg Glu Pro
145                 150                 155                 160

Gly Asn Asp Lys Phe Arg Arg Val Arg Leu Gly Asn Pro Arg Ile Lys
                165                 170                 175

Glu Ala Leu Ala Asp Arg Asp Gly Gly Val Glu Leu Leu Glu Ala Val
            180                 185                 190

Gly Phe Thr Val Gly Asp Glu Gly Gly Glu Pro Phe Ala Val Met Asp
        195                 200                 205

Glu Val Pro Ser Asp Pro Arg Leu Asn Gly Ile Arg Arg Ala Val Leu
    210                 215                 220

Leu Leu Glu Gly Ala His Pro Ser Ala Pro Pro Val Lys Ala Glu Ala
225                 230                 235                 240

Glu Ala Lys Glu Ser Cys Ser Asn Val Ser Asp Val Gln Glu Gly Ala
                245                 250                 255

Lys Thr Ile Asp Arg Gln Ile Arg Val Phe Val Ser Val Pro Gly Ser
            260                 265                 270

Ser Met Ala Gln Asn Asp Val Pro Asp Ser Phe Tyr Lys Leu Ser Gly
        275                 280                 285

Glu Glu Ile Arg Asn Glu Ala Lys Met Arg Arg Glu Arg Leu Glu Gln
```

```
    290                 295                 300

Ser Arg Leu Leu Ile Pro Lys Ser Tyr Lys Glu Lys Gln Ala Leu Ala
305                 310                 315                 320

Ala Arg Gln Lys Tyr Lys Gln Ala Val Ile Arg Val Gln Phe Pro Asp
                325                 330                 335

Arg Met Ile Leu Gln Gly Ile Phe Leu Pro Gly Glu Ala Thr Ser Ser
            340                 345                 350

Leu Tyr Glu Phe Val Thr Ser Ala Leu Lys Gln Ser Gly Leu Glu Phe
        355                 360                 365

Glu Leu Ile Ser Pro Ala Ile Pro Lys Pro Arg Val Val Pro His Phe
    370                 375                 380

Pro Asn Pro Gly Glu Arg Ala Arg Thr Leu Gln Glu Glu Glu Leu Val
385                 390                 395                 400

Pro Ser Ala Leu Leu Lys Phe Ile Pro Lys Glu Thr Asp Ser Met Val
                405                 410                 415

Phe Thr Gly Leu Leu Asp Glu Leu Leu Met Ala Ser Glu Pro Leu Pro
            420                 425                 430

Ala Ala Ser Gln
        435
```

<210> 5
<211> 3048
<212> DNA
<213> Artificial sequence

<220>
<223> expression cassette comprising GRUBX (1011-2390) operably linked to the prolamine promoter (1-654) and the T-Zein + T-Rubisco delta G terminator (2615-2808 and 2852-3048)

<400> 5

```
cttctacatc ggcttaggtg tagcaacacg actttattat tattattatt attattatta     60
ttattttaca aaaatataaa atagatcagt ccctcaccac aagtagagca agttggtgag    120
ttattgtaaa gttctacaaa gctaatttaa aagttattgc attaacttat ttcatattac    180
aaacaagagt gtcaatggaa caatgaaaac catatgacat actataattt tgtttttatt    240
attgaaatta tataattcaa agagaataaa tccacatagc cgtaaagttc tacatgtggt    300
gcattaccaa aatatatata gcttacaaaa catgacaagc ttagtttgaa aaattgcaat    360
ccttatcaca ttgacacata aagtgagtga tgagtcataa tattattttc tttgctaccc    420
atcatgtata tatgatagcc acaaagttac tttgatgatg atatcaaaga acatttttag    480
gtgcacctaa cagaatatcc aaataatatg actcacttag atcataatag agcatcaagt    540
aaaactaaca ctctaaagca accgatggga aagcatctat aaatagacaa gcacaatgaa    600
aatcctcatc atccttcacc acaattcaaa tattatagtt gaagcatagt agtaatttaa    660
atcaactagg gatatcacaa gtttgtacaa aaaagcaggc tggtaccggt ccggaattcc    720
cgggatatcg tcgacccacg cgtccgccaa tatcagattt ctttcatgaa ctccacttcc    780
aatttctcat tgcttcttct tcccatttcc acctccaaag ccatccttcc agaaaacctt    840
gttccttaca tttcttagcc ccaaaaaaga ttcccatctc aattccacaa aaaaacacaa    900
ggagatctaa ggaaattccc cgcctctata tatagagagg tggaattgtt cctgaatttg    960
gtttgaattg attgattgac agattttggt gagagggtgt tattgaaaaa atgggtgaca   1020
tgaaggataa agtcaaaggg ttcatgaaaa aagtcacatc ttcttcttca ggtaagttta   1080
aaggccaagg tagggttttg ggtggttcat cttcttcagg accctcaaat catgtcaata   1140
```

atttttcatc acatccccta aatacaaggc aagatcaaca accttcatat acaaaaactt 1200
cgcctcaaaa accaagtaat tctgatcaaa gaattgagaa tatatgtgaa attcagttca 1260
acaaaagtga atcaaaggat ggttttgatc catttggtga attagtcact tctgggaaga 1320
gaaacccaaa agggtattca cttactaatg tgtttgaatg ccctgtctgt ggtagtggtt 1380
ttgtttctga agaagaggtg tcaactcata ttgatagctg tttaagttct gaagtgtctt 1440
ctaatttggg agttgaaagt aaagttgaag ttaaaagtga attggaaaca tgtgttagtg 1500
catatgtttc agggaagccc tcagaagggt cagttgaagt ggtcattaag ttgttaaaga 1560
atattgtgaa ggaaccagag aatgccaagt ttaggaaaat aaggatgggg aatccaaaaa 1620
taaaaggtgc tataggtgat gttgtaggag gagtggagct attggaattt gttggatttg 1680
agttgaaaga agaaggtggg gaaatttggg ctgtgatgga tgttccttct gaagaacaac 1740
ttgttatgct taagaatgta gtttcactct tggaaccgaa gaaggttgaa gagttggcgt 1800
ccttatccca agttaaggcg agtgaaccag ttgagccgaa gaagattgat agacagattc 1860
gagtgttctt ttctgttccc gagagcgtag cagcaaaaat tgagctacct gattccttct 1920
ttaacctctc acgtgaggaa ttgagaagag aagcagagat gaggaagaag aaattagaag 1980
attccaaatt attgattcct aaatcttatc gggaaaagca ggcaaaagct gcaagaaaga 2040
agtacacaaa atccattatc cgtgtacagt ttccagatgg agcattgctt caaggtgtct 2100
ttctaccttc ggagccaact agtgctcttt atgagtttgt gagcgcagcg ttaaaggaac 2160
caagcttaga gttcgaattg ttacatccgg tgcttgttaa aaagcgggtg attccccatt 2220
ttccagctgc tggggagagg gctgtaacag ttgaagagga ggatttggtt cctgcagctc 2280
tactcaaatt taaacctatc gaaacagatt ctgttgtttt tactggtctt tgtaatgagc 2340
ttcttgaaat tagcgagccc ctcgagaccg gatcagttgc ttcctcgtaa gctctaaatt 2400
acatcagact ttgaattctt ctgagtgttg gaaaccttat aaaactctct gcgccgggaa 2460
tgctgcggcc gctctagagt atccctcgag gggcccaagc ttacgcgtac ccagctttct 2520
tgtacaaagt ggtgatatca caagcccggg cggtcttcta gggataacag ggtaattata 2580
tccctctaga tcacaagccc gggcggtctt ctacgatgat tgagtaataa tgtgtcacgc 2640
atcaccatgg gtggcagtgt cagtgtgagc aatgacctga atgaacaatt gaaatgaaaa 2700
gaaaaaaagt actccatctg ttccaaatta aaattcattt taacctttta ataggtttat 2760
acaataattg atatatgttt tctgtatatg tctaatttgt tatcatccgg gcggtcttct 2820
agggataaca gggtaattat atccctctag acaacacaca acaaataaga gaaaaaacaa 2880
ataatattaa tttgagaatg aacaaaagga ccatatcatt cattaactct tctccatcca 2940
tttccatttc acagttcgat agcgaaaacc gaataaaaaa cacagtaaat tacaagcaca 3000
acaaatggta caagaaaaac agttttccca atgccataat actcgaac 3048

<210> 6
<211> 1302
<212> DNA
<213> Oryza sativa

<400> 6
atgatgaagg aaaagatgaa ggatctcatg aggaaggtca cctcctcctc ctcctcctcc 60
tcgtcgtcct cctccttcaa gggcaccgcc cacgtcctcg gctccggccc cgacccctcc 120
tcccgccccc ccaaccctac ccctagtcgc cccgctgccc cccggcgaga ggccgccgcc 180
tccgcgaggc cgccctcctc cggcttcgcc ccctactccc cgctcatctc cacctcctcc 240
cgccgcaccg acccacccgc gggggcgggg gcgggggagg acgacgccgt cgcgtgcccc 300
agctgcgccg agccgttccc ctccgagctg gcggtgtcgg accacctcga cggctgcctc 360
gcggcggcgg ggggagcccg cccccgcgcg gccgcctacc tggccggcga ccccccccgcg 420
tccgccgtgg aggtggtgaa gaggctgctc gggaacctgc tctccgaccc ccggaacgac 480
aagtacagga aggtcaggct cgggaacccg aggatcaagg aggccctggc ggacagggag 540
ggcggggtgg atctcctcga ggccgtgggg ttcagggtcg ccgacgaggg cggggagctc 600
ttcgccctca tggacgaggt gcccggggac gcgaggctcg gcggcatcag gcaggccgtg 660
ctcctgctcg agagggcccg gccatcgacg ccgccgcaga cacaggcaga tgccaaagag 720
acttgcccga atggagttag cgaagagcag gggattaaga agccggttga tcgtcagatt 780
cgggtgttct tctctgttgc tgcaagttct gttgcagaaa atgatctacc agattctttc 840
tatagcctta gtaatgagga gatcaggaat gaggcaaaga tgaggaggga gaggctagaa 900
caatctcggt tgttgattcc aaagtcatat aaggagaagc aggcactggc tgcccgacag 960
aagtataaac aagctctgat tcgaattcag tttccggatg gagtaattct gcagggtgtg 1020
ttccttcccg cggagcccat tagttcacta tatgagtttg tcgcatcttc tttgaagcag 1080
ccaagtttgg aattcgatct tatctgtcca gcgggtccta ggacacgcgt catacctcct 1140
tttccaaaac caggggaaca ggcacgcaca ctgcgggatg aagacctagt cccttctgct 1200

cggctcacct tcaagcccaa agagactgat tcggttgtgt tcactggtct gttggacgag 1260
cttcttgaga ccagtgagcc gttcacatct gcatcctcat ga 1302

<210> 7
<211> 433
<212> PRT
<213> Oryza sativa

<400> 7

```
Met Met Lys Glu Lys Met Lys Asp Leu Met Arg Lys Val Thr Ser Ser
1               5                  10                  15

Ser Ser Ser Ser Ser Ser Ser Ser Ser Phe Lys Gly Thr Ala His Val
            20                  25                  30

Leu Gly Ser Gly Pro Asp Pro Ser Ser Arg Pro Ser Asn Pro Thr Pro
        35                  40                  45

Ser Arg Pro Ala Ala Pro Arg Arg Glu Ala Ala Ala Ser Ala Arg Pro
    50                  55                  60

Pro Ser Ser Gly Phe Ala Pro Tyr Ser Pro Leu Ile Ser Thr Ser Ser
65                  70                  75                  80

Arg Arg Thr Asp Pro Pro Ala Gly Ala Gly Ala Gly Glu Asp Asp Ala
                85                  90                  95

Val Ala Cys Pro Ser Cys Ala Glu Pro Phe Pro Ser Glu Leu Ala Val
            100                 105                 110

Ser Asp His Leu Asp Gly Cys Leu Ala Ala Ala Gly Gly Ala Arg Pro
        115                 120                 125

Arg Ala Ala Ala Tyr Leu Ala Gly Asp Pro Pro Ala Ser Ala Val Glu
    130                 135                 140

Val Val Lys Arg Leu Leu Gly Asn Leu Leu Ser Asp Pro Arg Asn Asp
145                 150                 155                 160

Lys Tyr Arg Lys Val Arg Leu Gly Asn Pro Arg Ile Lys Glu Ala Leu
                165                 170                 175

Ala Asp Arg Glu Gly Gly Val Asp Leu Leu Glu Ala Val Gly Phe Arg
            180                 185                 190

Val Ala Asp Glu Gly Gly Glu Leu Phe Ala Leu Met Asp Glu Val Pro
        195                 200                 205

Gly Asp Ala Arg Leu Gly Gly Ile Arg Gln Ala Val Leu Leu Leu Glu
    210                 215                 220

Arg Ala Arg Pro Ser Thr Pro Pro Gln Thr Gln Ala Asp Ala Lys Glu
225                 230                 235                 240

Thr Cys Pro Asn Gly Val Ser Glu Glu Gln Gly Ile Lys Lys Pro Val
                245                 250                 255

Asp Arg Gln Ile Arg Val Phe Phe Ser Val Ala Ala Ser Ser Val Ala
            260                 265                 270
```

```
Glu Asn Asp Leu Pro Asp Ser Phe Tyr Ser Leu Ser Asn Glu Glu Ile
        275                 280                 285

Arg Asn Glu Ala Lys Met Arg Arg Glu Arg Leu Glu Gln Ser Arg Leu
    290                 295                 300

Leu Ile Pro Lys Ser Tyr Lys Glu Lys Gln Ala Leu Ala Ala Arg Gln
305                 310                 315                 320

Lys Tyr Lys Gln Ala Leu Ile Arg Ile Gln Phe Pro Asp Gly Val Ile
                325                 330                 335

Leu Gln Gly Val Phe Leu Pro Ala Glu Pro Ile Ser Ser Leu Tyr Glu
            340                 345                 350

Phe Val Ala Ser Ser Leu Lys Gln Pro Ser Leu Glu Phe Asp Leu Ile
        355                 360                 365

Cys Pro Ala Gly Pro Arg Thr Arg Val Ile Pro Pro Phe Pro Lys Pro
    370                 375                 380

Gly Glu Gln Ala Arg Thr Leu Arg Asp Glu Asp Leu Val Pro Ser Ala
385                 390                 395                 400

Arg Leu Thr Phe Lys Pro Lys Glu Thr Asp Ser Val Val Phe Thr Gly
                405                 410                 415

Leu Leu Asp Glu Leu Leu Glu Thr Ser Glu Pro Phe Thr Ser Ala Ser
            420                 425                 430

Ser
```